DETERMINATION OF FOLDING PATHWAY SELECTION AND ORIGINS OF COOPERATIVITY OF NATURALLY OCCURRING AND DESIGNED CONSENSUS LEUCINE-RICH REPEAT PROTEINS

by

Thuy Phuong Dao

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

June 2014

Abstract

Thermodynamic and kinetic studies of linear repeat proteins have provided unique insights into cooperativity, detailed maps of local stabilities, and sequence determinants of folding pathways. Most of these studies have focused on α-helical repeat proteins. Additional work on repeat proteins that feature other types of secondary structures, such as β-strands, is essential for understanding different kinds of interactions found in much more complicated globular proteins. Here, we investigated the folding properties of a naturally occurring β-strand-containing leucine-rich repeat (LRR) protein PP32 as well as designed consensus bacterial LRR constructs.

PP32 contains five tandem LRRs flanked by α-helical and β-strand capping motifs on the N- and C-termini, respectively. Terminal caps are often observed in LRR proteins, but not in helical repeat proteins. Without the C-cap, PP32 is unfolded. Without the N-cap, PP32 is less stable, but retains its secondary structure. However, solution studies by NMR and mutational analysis show that removing the N-cap causes the first two repeats to exist in a molten-globule-like state, where secondary structures are formed but rigid tertiary packing is disrupted. Therefore, both caps are essential for structure formation and stability of PP32, though to different extents.

Although PP32 undergoes an equilibrium two-state unfolding transition, its kinetic folding mechanism is more complicated, with the formation of an on-pathway intermediate as the rate-limiting step. Φ-value analysis reveals a

highly polarized transition state involving repeat 5 and part of repeat 4. Hydrogen exchange monitored by NMR spectroscopy shows that PP32 is most stable towards the C-terminus. Therefore, the folding pathway for PP32 is dictated by local stability, as observed for α-helical repeat proteins.

Whereas the studies of naturally occurring repeat proteins permit the variations in repeat sequence to be investigated, those of designed consensus repeats proteins, in which the repeats are identical (or nearly identical), allow us to resolve the intrinsic energy of individual repeats and interfacial energy between neighboring repeats. We have been able to create consensus bacterial LRR constructs that are well-behaved, stable and unfold via cooperative transitions. By fitting a nearest-neighbor Ising model to the unfolding transitions, we have determined that folding of individual repeats is unfavorable but the interactions between adjacent repeats are highly favorable, consistent with the high cooperativity observed for LRR proteins.

Thesis Advisor:	Doug Barrick, Ph.D.
Second Reader:	Vincent J. Hilser, Ph.D.
Thesis Committee:	Bertrand Garcia-Moreno E., Ph.D.
	Juliette T. J. Lecomte, Ph.D.
	Joel R. Tolman, Ph.D.

To my uncle Thuyết and his loving wife Ngọc Lan

Acknowledgments

I would like to thank my thesis advisor, Dr. Doug Barrick, for all of his wonderful guidance, encouragement and patience over the years. His love for science, constant excitement for data, any data, and thoughtful analysis of results inspire me to work hard to become a better scientist. Aside from repeat proteins, I will miss our morning chats about soccer, football, our tomatoes, Doug's Japanese rice bowls, and much more!!

I would like to thank my thesis committee members, Drs. Bertrand Garcia-Moreno E., Vincent Hilser, Juliette Lecomte and Joel Tolman for the challenging questions, and the very useful comments and suggestions, especially early on in my graduate school years. They helped me tremendously in thinking critically and presenting my work logically.

I would like to thank Dr. Ananya Majumdar for all of his help with NMR, from training to experimental setups, and his friendship. It is always fun to hear his many interesting life stories.

I would like to thank all of the past and present members of the Barrick lab for being a wonderful support system and for putting up with me. Christine Hatem has been my bay mate for as long as I can remember. She is very kind, understanding and generous. She listens to all of my problems and is always there for me. I really appreciate having her in my life. Dr. Ellen Kloss took me on when I was a recruit and helped me find my way around the lab, the instruments, and the proteins. I am very glad to have known her as a scientist and a friend. Dr.

Andrea Carter loves and excels at explaining scientific concepts. I learnt a lot from her and very much enjoy our random soccer texts. Dr. Scott Johnson is always funny in his own sarcastic way, which makes great early-morning lab conversations. I also love his long emails about the police academy and patrols. Jacob Marold was my first rotation student and I could not have asked for a better one. He is devoted, meticulous, thoughtful and I have learned much more from him than he has from me. Jake is also extremely considerate and caring, so I am very glad he is also my friend! Even though I have known Dr. Katie Tripp for less than two years, she has been my sounding board for most of that time, making the thesis-writing process much more bearable.

I would like to thank my classmates, especially Matt Preimesberger (and his sweet girlfriend Carla), Helen Jun, Jackson Buss, Gustavo Afanador-Gonzalez, and the pseudo classmate Matt Pond (and his wonderful wife Monique). We have had many food adventures and long, fun nights that might or might not involve crazy, violent Thuy.

I would like to thank Ranice Crosby for taking care of us at all times. I would have missed so many deadlines and been very lost without her help.

I would like to thank the office staff, especially Jerry Levin, Jess (soon-to-be Appel) Bailey, Ken Rutledge and Lexie Ebert for making sure research can be carried out smoothly and for entertaining me when I bug them. A special shout-out to Jess who quickly became one of my dearest friends. I will always

remember the fun and the stressful Paco moments we shared. I am very excited to be celebrating Jess and her sweetheart Ben on their special day!!

I would like to thank my Whitman professors, Drs. Daniel Vernon and Doug Juers, and Nancy Forsthoefel for providing me the opportunities to discover how fun research is. Doug's passion for teaching and biophysics inspired me to pursue the subject for my graduate work.

I would like to thank my family for believing in me and always being supportive of my decisions. I would especially like to thank my uncle Thuyết and my aunt Ngọc Lan who have been raising me as their own child. My uncle is wise, stern but fair. My aunt is gentle and kind-hearted. Together, they have taught me to be caring, considerate, hard-working and persistent. I hope to eventually make a difference in someone's life as they have in mine.

Last but not least, I would like to thank Carlos for being there for me during the last few years. He is always patient, encouraging, and willing to help me with whatever I need, from NMR, to house work, to chauffeuring, regardless of how busy he is. I can't wait for our many journeys ahead (including him being my boss…yikes!!)

Table of Contents

Abstract ii

Acknowledgements v

Table of Contents viii

List of Tables x

List of Figures xi

Chapter I. Introduction 1

1.1 Repeat proteins are ideal for exploring folding pathways and cooperativity 2

1.2 Most repeat proteins unfold through equilibrium two-state transitions 4

1.3 Folding kinetics of repeat proteins 6

1.4 Overview of LRR proteins 7

1.5 References 12

1.6 Appendix 19

Chapter II. Capping motifs stabilize the LRR protein PP32 and rigidify adjacent repeats (This chapter is reprinted with permission from the authors. T.P. Dao, A. Majumdar and D. Barrick (2014). *Protein Science* 23, 801-11.) **20**

2.1 Abstract 21

2.2 Introduction 22

2.3 Results 25

2.4 Discussion 33

2.5 Materials and Methods 37

2.6 References 40

Chapter III. The highly polarized C-terminal transition state of the leucine-rich repeat domain of PP32 is governed by local stability 53

3.1 Abstract 54

3.2 Introduction 55

3.3 Results 58

3.4 Discussion 67

3.5 Materials and Methods 74

3.6 References 79

Chapter IV. An Ising analysis of consensus LRR protein folding: high cooperativity from very strong interfacial interactions 95

4.1 Abstract 96

4.2 Introduction 97

4.3 Results 100

4.4 Discussion 106

4.5 Conclusion 110

4.6 Materials and Methods 112

4.7 References 115

Lists of Tables

Chapter II

Table 2.1 Thermodynamic unfolding parameters for hAnp32 1-149, hAnp32A 1-154 and cap variants 44

Table 2.2 Thermodynamic unfolding parameters for core packing substitutes in full-length and Δ_{NCap} PP32 45

Chapter III

Table 3.1 Fitted parameters for two-state equilibrium unfolding and three-state kinetic refolding and unfolding PP32 variants 85

Chapter IV

Table 4.1 Thermodynamic unfolding parameters of different consensus LRR constructs 118

Table 4.2 Parameters obtained from fitting of the Ising model to unfolding transitions of consensus LRR constructs 119

Lists of Figures

Chapter I

Figure 1.1 Diversity in structures of repeat proteins. 16

Figure 1.2 Diversity in structures of LRR proteins in different subfamilies 17

Figure 1.3 Hidden Markov models of representative LRR subfamilies 18

Figure A-1 Structure and urea-induced denaturation of Rna1p 19

Chapter II

Figure 2.1 Structure and sequence of PP32 46

Figure 2.2 CD spectra and equilibrium unfolding of hAnp32A with different C-termini 47

Figure 2.3 Solution spectroscopy and equilibrium unfolding of PP32 and PP32 Y131F/D146L 48

Figure 2.4 Equilibrium unfolding and NMR spectroscopy of PP32 and Δ_{NCap} PP32 49

Figure 2.5 RDC of Δ_{Ncap} PP32 50

Figure 2.6 The ^{15}N spin relaxation parameters of Δ_{NCap} PP32 51

Figure 2.7 CD spectroscopy and equilibrium unfolding of core packing substitutions in PP32 and Δ_{NCap} PP32 52

Chapter III

Figure 3.1 Structure and sequence of PP32 86

Figure 3.2 Solution spectroscopy and equilibrium unfolding of PP32 87

Figure 3.3 Refolding and unfolding kinetics of PP32 88

Figure 3.4 Urea-induced equilibrium unfolding of PP32 variants 89

Figure 3.5 Chevron plots for variants of PP32 90

Figure 3.6 Φ-values and the folding pathway of PP32 91

Figure 3.7 Residue-specific hydrogen exchange of PP32 92

Figure 3.8 Folding pathway of PP32 93

Figure 3.9 Effects of destabilizing the C-terminus to the folding pathway of PP32 94

Chapter IV

Figure 4.1 Design of consensus bacterial LRR protein repeats 120

Figure 4.2 Concentration dependence of sedimentation and urea-induced denaturation of consensus LRR constructs 122

Figure 4.3 Far-UV scans and equilibrium unfolding of different consensus LRR constructs 123

Figure 4.4 Ising analysis of unfolding transitions of consensus LRR constructs 124

Chapter I

Introduction

1.1 Repeat proteins are ideal for exploring folding pathways and cooperativity

The folding of polypeptides into well-defined, functional “native” state is arguably the most fundamental biological process in the cell. In most cases, the native conformation of a protein is solely defined by its amino acid sequence, as first described for ribonuclease A (Anfinsen, 1973). The folding process normally involves the spontaneous organization of a chain of a hundred or more amino acids into a unique structure involving hundreds of narrowly defined backbone and side chain dihedral angles. Owing to the vast number of conformations accessible to an unfolded polypeptide, it is remarkable that proteins fold at all, let alone so quickly, sometimes within sub-millisecond timescales, as observed for many small proteins (Kubelka et al., 2004; Yang and Gruebele, 2003). Moreover, the folding reaction is often cooperative, involving only unfolded and folded species, but not partially folded species. Although an extensive amount of work has been done, the origin of cooperativity is still not fully understood (Sosnick and Barrick, 2011).

Cooperativity can be quantified by the energy distribution within a protein (Aksel and Barrick, 2009). Dissecting the contribution of a region to the folding process typically requires the protein to be studied with that region removed. For globular proteins, which comprise many contacts between regions that are distant in sequence, removing a region would most likely disrupt the overall fold, preventing interpretation of any available results. In

favorable cases, this problem can be bypassed by using hydrogen exchange (HX) methods, which measure stability of the native protein at single-residue resolution (Hvidt and Linderstrøm-Lang, 1954; Krishna et al., 2004). For a complete and detailed stability map, the HX experiments have to be carried out over a range of mildly destabilizing conditions. This process is not only labor-intensive and time-consuming, but also limited to proteins of high global stability. Moreover, the irregular tertiary structures of globular proteins also make it hard to compare different regions.

In contrast to globular proteins, repeat proteins are simpler and more regular in architecture, and lack sequence-distant contacts. These proteins contain repeated units of highly similar secondary structure that stack together in a linear array. The regular and modular nature of repeat proteins allows the contribution of individual repeats to be dissected and compared by adding and deleting repeats (Kloss and Barrick, 2009; Mello and Barrick, 2004; Tripp and Barrick, 2004; Tsytlonok et al., 2013a; Vieux and Barrick, 2011). Moreover, the simple, repetitive organization of secondary and tertiary structure can be extended to primary structure through consensus design, further simplifying analysis and comparison (Aksel and Barrick, 2009; Aksel et al., 2011; Main et al., 2003; Parker et al., 2014; Stumpp et al., 2003; Tripp and Barrick, 2007; Wetzel et al., 2008)

Though simple in architecture, different types of repeat proteins are highly diverse in structure, with repeats containing mainly helices or only coils,

or mixtures of β-strands and other secondary structure elements (Figures 1.1 and 1.2). The different types of repeats lead to a diversity of cellular functions, with scaffolding serving as a common role (Andrade et al., 2001). From the perspective of folding studies, this diversity in secondary structures allows for the understanding of different simpler structural units, which can then be applied to much more complicated globular proteins.

Folding studies of repeat proteins have focused primarily on α-helical ankyrin repeat and tetratricopeptide repeat (TPR) and to a lesser extent on β-strand-containing leucine-rich repeats (LRR). Owing to the differences in structures of these two types of repeats, the average change in surface area upon folding of individual α-helical repeats is about twice that of β-stranded repeats (Kloss et al., 2008). Although the inter-repeat interactions are different for the two types of repeat proteins (helix packing vs. β-sheet formation), the average surface area buried between adjacent repeats is similar. How do the folding processes for α-helical and β-strand containing repeat proteins differ? What are the consequence of the structural differences on folding? Even though much is known about the folding properties of these proteins, especially α-helical repeat proteins, more work is needed as described below.

1.2 Most repeat proteins unfold through equilibrium two-state transitions

Due to the lack of contacts between regions far apart in sequence, elongated, modular repeat proteins might be expected to unfold via multiple

transitions. However, most proteins studied to date are highly cooperative in folding, including α-helical repeat proteins (Aksel et al., 2011; Lowe and Itzhaki, 2007; Main et al., 2003; Mosavi et al., 2002; Tang et al., 1999; Zweifel and Barrick, 2001), LRR proteins (Courtemanche and Barrick, 2008a; Kelly et al., 2014; Kloss and Barrick, 2008), and coiled pentapeptide repeat protein HetL (Dao and Barrick, unpublished data). Moreover, the limit of cooperativity appears to be higher in repeat proteins than in globular proteins, which tend to unfold via multiple transitions at longer chain lengths. For LRR proteins, the observed *m*-values, which indicate the size of the cooperative units, are higher than predicted based on empirical values from studies of globular proteins (Myers et al., 1995), further supporting high cooperativity in folding of repeat proteins.

However, multiple-state equilibrium unfolding has also been observed for some repeat proteins (Junker et al., 2006; Kamen et al., 2000; Tsytlonok et al., 2013b; Werbeck and Itzhaki, 2007; Zeeb et al., 2002). Some of these proteins are very large, containing up to 20 repeats and over 500 residues. For cooperativity to be observed, thermodynamic coupling from one end to the other is needed, which can be hard for very long modular proteins. It has also been shown that highly skewing the stability distribution across a repeat domain without enhancing inter-repeat stability can disrupt the two-state folding mechanism (Bradley and Barrick, 2002; Street et al., 2007; Tripp and Barrick, 2007).

High cooperativity is also observed for equilibrium unfolding of designed consensus α-helical ankyrin repeats and to a lesser extent, for α-helical TPR proteins. One-dimensional Ising analysis of the Notch ankyrin domain and consensus α-helical repeat proteins have shown that folding of single repeats is unfavorable, whereas interfacial interactions between adjacent repeats are highly favorable (Aksel and Barrick, 2009; Aksel et al., 2011; Kajander et al., 2005; Mello and Barrick, 2004; Wetzel et al., 2008). This ensures coupled folding of individual repeats which cannot be folded on their own. Therefore, the cooperativity observed in the folding of repeat proteins is the result of: 1) a fairly uniform energetic distribution, 2) highly stabilizing interfacial energies, and 3) unfavorable intrinsic folding energies.

1.3 Folding kinetics of repeat proteins

The cooperativity in equilibrium unfolding is not high enough to drive two-state kinetic folding mechanism in repeat proteins. To date, only one protein, the LRR domain of InlB, exhibits simple two-state kinetics (Courtemanche and Barrick, 2008a). For most proteins, both α-helical and β-stranded, more than one kinetic phase is observed as well as a non-linear chevron plot (Kelly et al., 2014; Kloss and Barrick, 2008; Löw et al., 2007; Lowe and Itzhaki, 2007b; Mello et al., 2005; Tang et al., 1999; Tsytlonok et al., 2013b), with the most common mechanism consisting of an on-pathway intermediate that separates the denatured and native states.

Φ-value analysis on repeat proteins that unfold through equilibrium two-state transitions show that these proteins fold along well-defined pathways that typically involve around half of the repeats (Bradley and Barrick, 2006; Courtemanche and Barrick, 2008b; Tang et al., 2003). Considering that multiple parallel pathways might be expected for these proteins due to the high redundancy in the repeat structures, the observed polarized pathways are somewhat surprising. For the seven-repeat Notch ankyrin domain, the centralized folding route corresponds to one of the two low energy routes on the experimentally determined energy landscape (Bradley and Barrick, 2006; Mello and Barrick, 2004). More strikingly, when the last two repeats were replaced with consensus repeats, the region of high stability is shifted towards the C-terminus, as is the transition state (Tripp and Barrick, 2007, 2008). For consensus ankyrin repeat constructs of nearly identical repeat sequences, multiple parallel pathways are observed, because these pathways have the same energy (Aksel and Barrick, in press). These results suggest that, for α-helical repeat proteins, folding pathways are dictated by local stability.

1.4 Overview of LRR proteins

Studies of α-helical repeat proteins have provided great insights into the origin of cooperativity as well as kinetic pathway selection during folding. However, much less is known about β-strand-containing repeat proteins, which have been less amenable to whole-repeat manipulations and

consensus design than their α-helical counterparts. One of the most attractive targets for filling in these gaps in our understanding of β-strand repeat proteins is the leucine-rich repeat (LRR) family of repeat proteins. Below I describe the structural features of LRR proteins, as well as some structural and thermodynamic features of LRR proteins that complicate folding studies and limit analysis.

LRR proteins contain 2-45 tandem repeats; each LRR is 20-30 amino acids in length, depending on the subfamily (see below). LRRs contain the conserved sequence LxxLxLxxN/CxL (Enkhbayar et al., 2004). The defining structural feature of each LRR is the existence of a β-strand (Figure 1.2). When the repeats stack together, the β-strands form a β-sheet that lines up the concave surface of these proteins. LRRs are divided into seven subfamilies that differ in repeat lengths and have additional family-specific sequence features. Depending on length, different secondary structures are observed on the convex surface, ranging from α-helix for 28-30-residue long RI-type repeats (Figure 1.2A) to extended structure for 20-residue long repeats of the bacterial subfamily (Figure 1.2B), with 3_{10} and PII helices or extended loops for the repeats with intermediate lengths from subfamilies such as typical and SDS22-like (Figures 1.2C and 1.2D).

Another difference between long (RI) and short LRRs is that the LRR proteins with shorter repeats typically have structurally distinct capping motifs, on either or both ends. For the two LRRs examined so far (InlB and YopM),

removing these caps leads to partial or complete unfolding of the LRR domains (Courtemanche and Barrick, unpublished observations; Kloss and Barrick, 2009; Kloss and Barrick, unpublished observations). The caps may be essential for constraining the repeats and protecting the hydrophobic core of the proteins from being exposed to solvent. Since it has not been possible so far to study the folding of LRR proteins without their caps, the local stabilities of different regions of LRR proteins (i.e. their energy landscape) has not been elucidated. Kinetic folding pathways for the two LRRs have been determined (Courtemanche and Barrick, 2008b; Kelly et al., 2014), but without explicitly determined energy landscapes, the origin of these pathways remains unknown.

It has also been challenging to generate and study consensus LRR proteins, compared to their α-helical counterparts. To date, the folding of two sets of consensus LRR constructs have been reported. In contrast to the highly cooperative unfolding of naturally occurring LRR proteins and consensus α-helical repeat proteins, consensus LRR constructs unfold at low denaturant concentrations, through shallow transitions, with much lower than predicted *m*-values (Parker et al., 2014; Stumpp et al., 2003). Both of these studies focused on the RI-like subfamily of LRR proteins, which contain long repeats of $\beta\alpha$ secondary structures, and lack capping motifs. The naturally occurring proteins from this family have not been studied. It could be that the lack of cooperativity observed for the folding of these constructs is intrinsic to

this subfamily and not due to consensus design, as supported by preliminary result for a naturally occurring RI-like protein Rna1p (see Appendix A).

Sequence alignment shows higher conservation at most hydrophobic positions for other subfamilies compared to the RI-like subfamily (Figure 1.2). Studies of consensus constructs from another LRR subfamily would be informative. However, as mentioned above, studies of consensus LRRs of shorter lengths are stymied by the likely requirement for capping motifs. The compatibility of the capping motifs to the surface of a consensus LRR might be difficult to design.

Additional work on LRR proteins will extend our understanding to folding of repeats of different secondary structures, which will provide insight into the folding of more complicated globular proteins. Are folding pathways selected based on local stability? Is folding of individual secondary structural units always unfavorable? Are the interactions between structural units always favorable?

My thesis work focused on the folding of the LRR domain of human tumor suppressor protein PP32, which contains five LRR repeats flanked by N- and C-terminal capping motifs. In addition, I have successfully designed and studied consensus LRR constructs from the bacterial subfamily. In Chapter II, the roles of the capping motifs of PP32 are explored through a combination of structural studies by NMR spectroscopy, equilibrium thermodynamic measurements of folding, and mutational analysis. Chapter III

describes the elucidation of the folding pathway of PP32 and how it is selected, using native-state HX to determine the stability distribution and compare with the folding pathway. Chapter IV focuses on the design and construction of stable consensus LRR constructs that undergo highly cooperative unfolding transitions, and the determination of interfacial and intrinsic energies using a nearest-neighbor Ising model.

1.5 References

Aksel, T., and Barrick, D. (2009). Chapter 4 Analysis of Repeat-Protein Folding Using Nearest-Neighbor Statistical Mechanical Models. In Methods in Enzymology, J.M.H. and G.K.A. Michael L. Johnson, ed. (Academic Press), pp. 95–125.

Aksel, T., Majumdar, A., and Barrick, D. (2011). The Contribution of Entropy, Enthalpy, and Hydrophobic Desolvation to Cooperativity in Repeat-Protein Folding. Structure *19*, 349–360.

Andrade, M.A., Perez-Iratxeta, C., and Ponting, C.P. (2001). Protein Repeats: Structures, Functions, and Evolution. J. Struct. Biol. *134*, 117–131.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science *181*, 223–230.

Bradley, C.M., and Barrick, D. (2002). Limits of Cooperativity in a Structurally Modular Protein: Response of the Notch Ankyrin Domain to Analogous Alanine Substitutions in Each Repeat. J. Mol. Biol. *324*, 373–386.

Bradley, C.M., and Barrick, D. (2006). The Notch Ankyrin Domain Folds via a Discrete, Centralized Pathway. Structure *14*, 1303–1312.

Courtemanche, N., and Barrick, D. (2008a). Folding thermodynamics and kinetics of the leucine-rich repeat domain of the virulence factor Internalin B. Protein Sci. Publ. Protein Soc. *17*, 43–53.

Courtemanche, N., and Barrick, D. (2008b). The leucine-rich repeat domain of Internalin B folds along a polarized N-terminal pathway. Struct. Lond. Engl. 1993 *16*, 705–714.

Enkhbayar, P., Kamiya, M., Osaki, M., Matsumoto, T., and Matsushima, N. (2004). Structural principles of leucine-rich repeat (LRR) proteins. Proteins Struct. Funct. Bioinforma. *54*, 394–403.

Hvidt, A., and Linderstrøm-Lang, K. (1954). Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. Biochim. Biophys. Acta *14*, 574–575.

Junker, M., Schuster, C.C., McDonnell, A.V., Sorg, K.A., Finn, M.C., Berger, B., and Clark, P.L. (2006). Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. Proc. Natl. Acad. Sci. U. S. A. *103*, 4918–4923.

Kajander, T., Cortajarena, A.L., Main, E.R.G., Mochrie, S.G.J., and Regan, L. (2005). A New Folding Paradigm for Repeat Proteins. J. Am. Chem. Soc. *127*, 10188–10190.

Kamen, D.E., Griko, Y., and Woody, R.W. (2000). The stability, structural organization, and denaturation of pectate lyase C, a parallel beta-helix protein. Biochemistry (Mosc.) *39*, 15932–15943.

Kelly, S.E., Meisl, G., Rowling, P.J.E., McLaughlin, S.H., Knowles, T., and Itzhaki, L.S. (2014). Diffuse transition state structure for the unfolding of a leucine-rich repeat protein. Phys. Chem. Chem. Phys.

Kloss, E., and Barrick, D. (2008). Thermodynamics, Kinetics, and Salt dependence of Folding of YopM, a Large Leucine-rich Repeat Protein. J. Mol. Biol. *383*, 1195–1209.

Kloss, E., and Barrick, D. (2009). C-terminal deletion of leucine-rich repeats from YopM reveals a heterogeneous distribution of stability in a cooperatively folded protein. Protein Sci. Publ. Protein Soc. *18*, 1948–1960.

Kloss, E., Courtemanche, N., and Barrick, D. (2008). Repeat-protein folding: New insights into origins of cooperativity, stability, and topology. Arch. Biochem. Biophys. *469*, 83–99.

Krishna, M.M.G., Hoang, L., Lin, Y., and Englander, S.W. (2004). Hydrogen exchange methods to study protein folding. Methods *34*, 51–64.

Kubelka, J., Hofrichter, J., and Eaton, W.A. (2004). The protein folding “speed limit.”Curr. Opin. Struct. Biol. *14*, 76–88.

Löw, C., Weininger, U., Zeeb, M., Zhang, W., Laue, E.D., Schmid, F.X., and Balbach, J. (2007). Folding mechanism of an ankyrin repeat protein: scaffold and active site formation of human CDK inhibitor p19(INK4d). J. Mol. Biol. *373*, 219–231.

Lowe, A.R., and Itzhaki, L.S. (2007a). Biophysical Characterisation of the Small Ankyrin Repeat Protein Myotrophin. J. Mol. Biol. *365*, 1245–1255.

Lowe, A.R., and Itzhaki, L.S. (2007b). Rational redesign of the folding pathway of a modular protein. Proc. Natl. Acad. Sci. *104*, 2679–2684.

Main, E.R.G., Xiong, Y., Cocco, M.J., D’Andrea, L., and Regan, L. (2003). Design of Stable α-Helical Arrays from an Idealized TPR Motif. Structure *11*, 497–508.

Mello, C.C., and Barrick, D. (2004). An experimentally determined protein folding energy landscape. Proc. Natl. Acad. Sci. U. S. A. *101*, 14102–14107.

Mello, C.C., Bradley, C.M., Tripp, K.W., and Barrick, D. (2005). Experimental Characterization of the Folding Kinetics of the Notch Ankyrin Domain. J. Mol. Biol. *352*, 266–281.

Mosavi, L.K., Minor, D.L., and Peng, Z. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. Proc. Natl. Acad. Sci. *99*, 16029–16034.

Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci. Publ. Protein Soc. *4*, 2138–2148.

Parker, R., Mercedes-Camacho, A., and Grove, T.Z. (2014). Consensus design of a NOD receptor leucine rich repeat domain with binding affinity for a muramyl dipeptide, a bacterial cell wall fragment. Protein Sci. n/a–n/a.

Sosnick, T.R., and Barrick, D. (2011). The folding of single domain proteins — have we reached a consensus? Curr. Opin. Struct. Biol. *21*, 12–24.

Street, T.O., Bradley, C.M., and Barrick, D. (2007). Predicting coupling limits from an experimentally determined energy landscape. Proc. Natl. Acad. Sci. *104*, 4907–4912.

Stumpp, M.T., Forrer, P., Binz, H.K., and Plückthun, A. (2003). Designing Repeat Proteins: Modular Leucine-rich Repeat Protein Libraries Based on the Mammalian Ribonuclease Inhibitor Family. J. Mol. Biol. *332*, 471–487.

Tang, K.S., Guralnick, B.J., Wang, W.K., Fersht, A.R., and Itzhaki, L.S. (1999). Stability and folding of the tumour suppressor protein p16. J. Mol. Biol. *285*, 1869–1886.

Tang, K.S., Fersht, A.R., and Itzhaki, L.S. (2003). Sequential Unfolding of Ankyrin Repeats in Tumor Suppressor p16. Structure *11*, 67–73.

Tripp, K.W., and Barrick, D. (2004). The Tolerance of a Modular Protein to Duplication and Deletion of Internal Repeats. J. Mol. Biol. *344*, 169–178.

Tripp, K.W., and Barrick, D. (2007). Enhancing the Stability and Folding Rate of a Repeat Protein through the Addition of Consensus Repeats. J. Mol. Biol. *365*, 1187–1200.

Tripp, K.W., and Barrick, D. (2008). Rerouting the folding pathway of the Notch ankyrin domain by reshaping the energy landscape. J. Am. Chem. Soc. *130*, 5681–5688.

Tsytlonok, M., Sormanni, P., Rowling, P.J.E., Vendruscolo, M., and Itzhaki, L.S. (2013a). Subdomain Architecture and Stability of a Giant Repeat Protein. J. Phys. Chem. B *117*, 13029–13037.

Tsytlonok, M., Craig, P.O., Sivertsson, E., Serquera, D., Perrett, S., Best, R.B., Wolynes, P.G., and Itzhaki, L.S. (2013b). Complex Energy Landscape of a Giant Repeat Protein. Structure *21*, 1954–1965.

Vieux, E.F., and Barrick, D. (2011). Deletion of internal structured repeats increases the stability of a leucine-rich repeat protein, YopM. Biophys. Chem. *159*, 152–161.

Werbeck, N.D., and Itzhaki, L.S. (2007). Probing a moving target with a plastic unfolding intermediate of an ankyrin-repeat protein. Proc. Natl. Acad. Sci. *104*, 7863–7868.

Wetzel, S.K., Settanni, G., Kenig, M., Binz, H.K., and Plückthun, A. (2008). Folding and Unfolding Mechanism of Highly Stable Full-Consensus Ankyrin Repeat Proteins. J. Mol. Biol. *376*, 241–257.

Yang, W.Y., and Gruebele, M. (2003). Folding at the speed limit. Nature *423*, 193–197.

Zeeb, M., Rösner, H., Zeslawski, W., Canet, D., Holak, T.A., and Balbach, J. (2002). Protein folding and stability of human CDK inhibitor p19INK4d. J. Mol. Biol. *315*, 447–457.

Zweifel, M.E., and Barrick, D. (2001). Studies of the Ankyrin Repeats of the Drosophila melanogaster Notch Receptor. 2. Solution Stability and Cooperativity of Unfolding†. Biochemistry (Mosc.) *40*, 14357–14367.

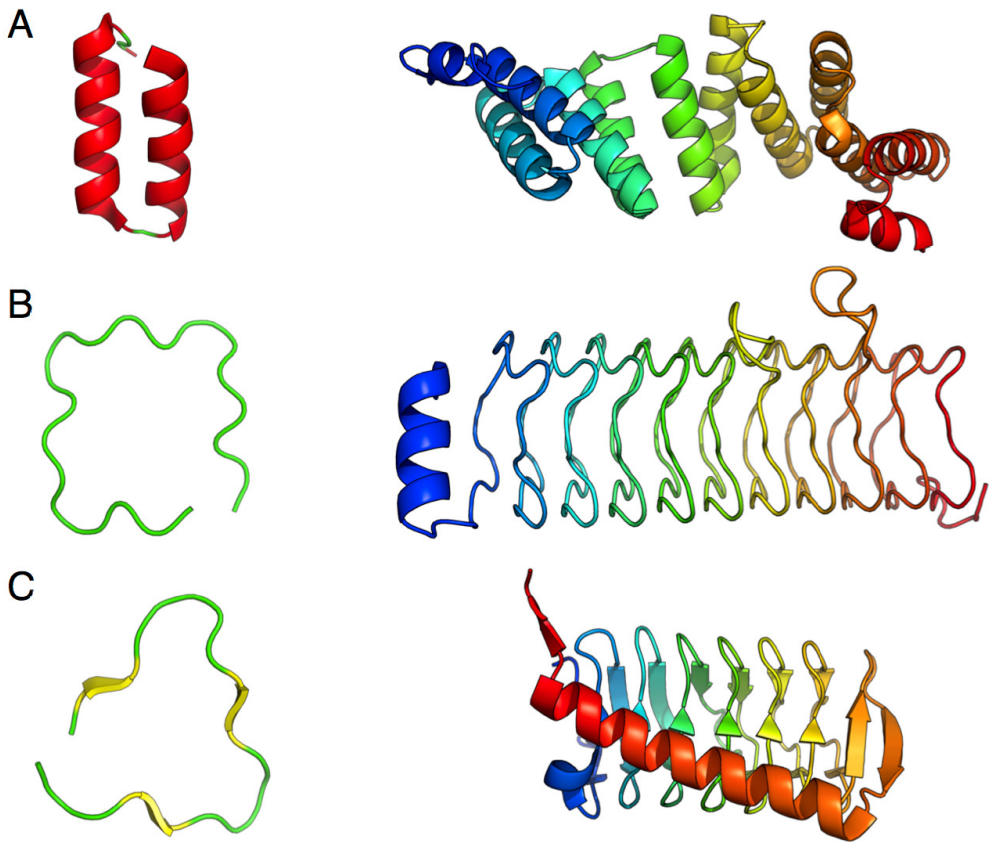


Figure 1.1. Diversity in structures of repeat proteins. (A) The crystal structure of tetratricopeptide repeat protein YrrB (PDB: 2Q7F). (B) The crystal structure of pentapeptide repeat protein HetL (PDB: 3DU1). (C) The crystal structure of hexapeptide repeat protein RicA (PDB: 4N27). Left: secondary structure of individual repeats. α-helices are in red; β-strands are in yellow; coils are in green. Right: repeats stack together to form linear arrays, colored in rainbow, with the N-terminus in blue and the C-terminus in red.

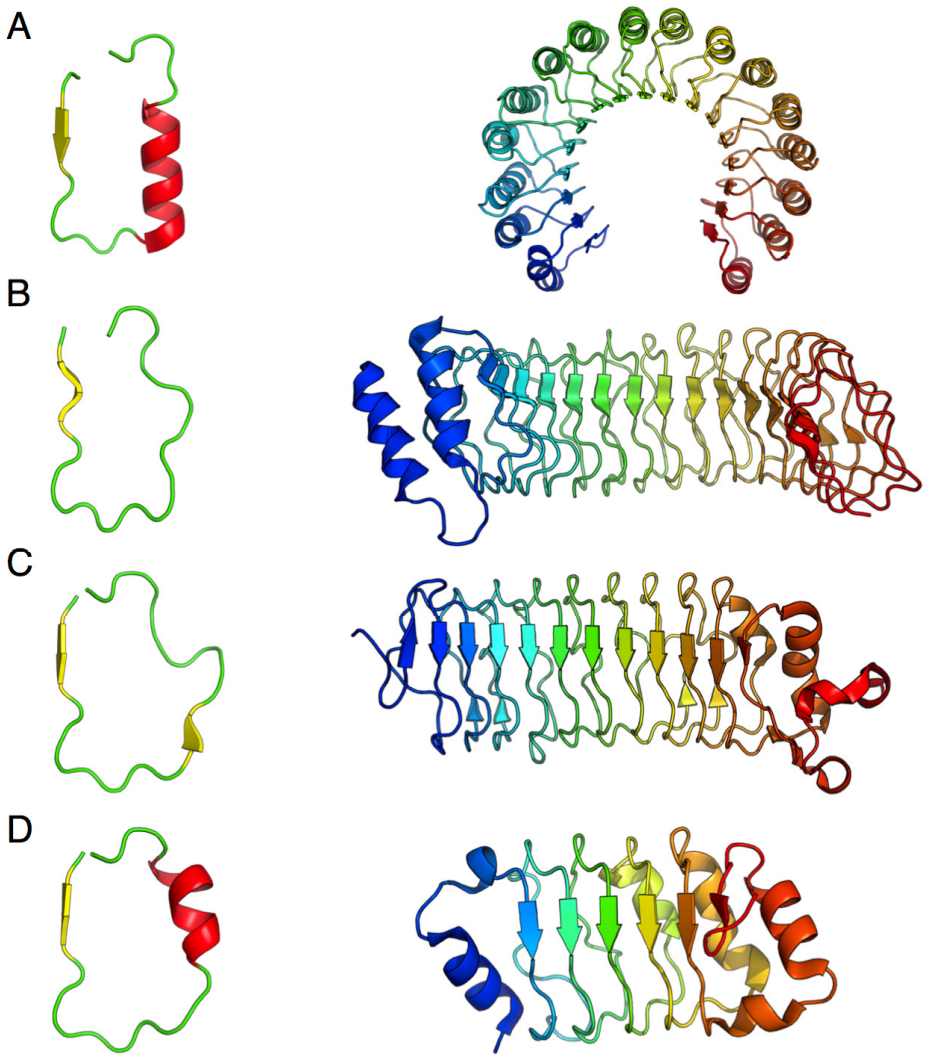


Figure 1.2. Diversity in structures of LRR proteins in different subfamilies. (A) The structure of LRR protein ribonuclease inhibitor (PDB: 2BNH). (B) The structure of bacterial LRR protein YopM (PDB: 1JL5). (C) The structure of typical LRR protein nogo-receptor-2 (PDB: 4P8S). (D) The structure of SDS22-like LRR protein PP32 (PDB: 2JE1). Left: secondary structure of individual repeats. α-helices are in red; β-strands are in yellow; coils are in green. Right: repeats stack together to form linear arrays, colored in rainbow, with the N-terminus in blue and the C-terminus in red.

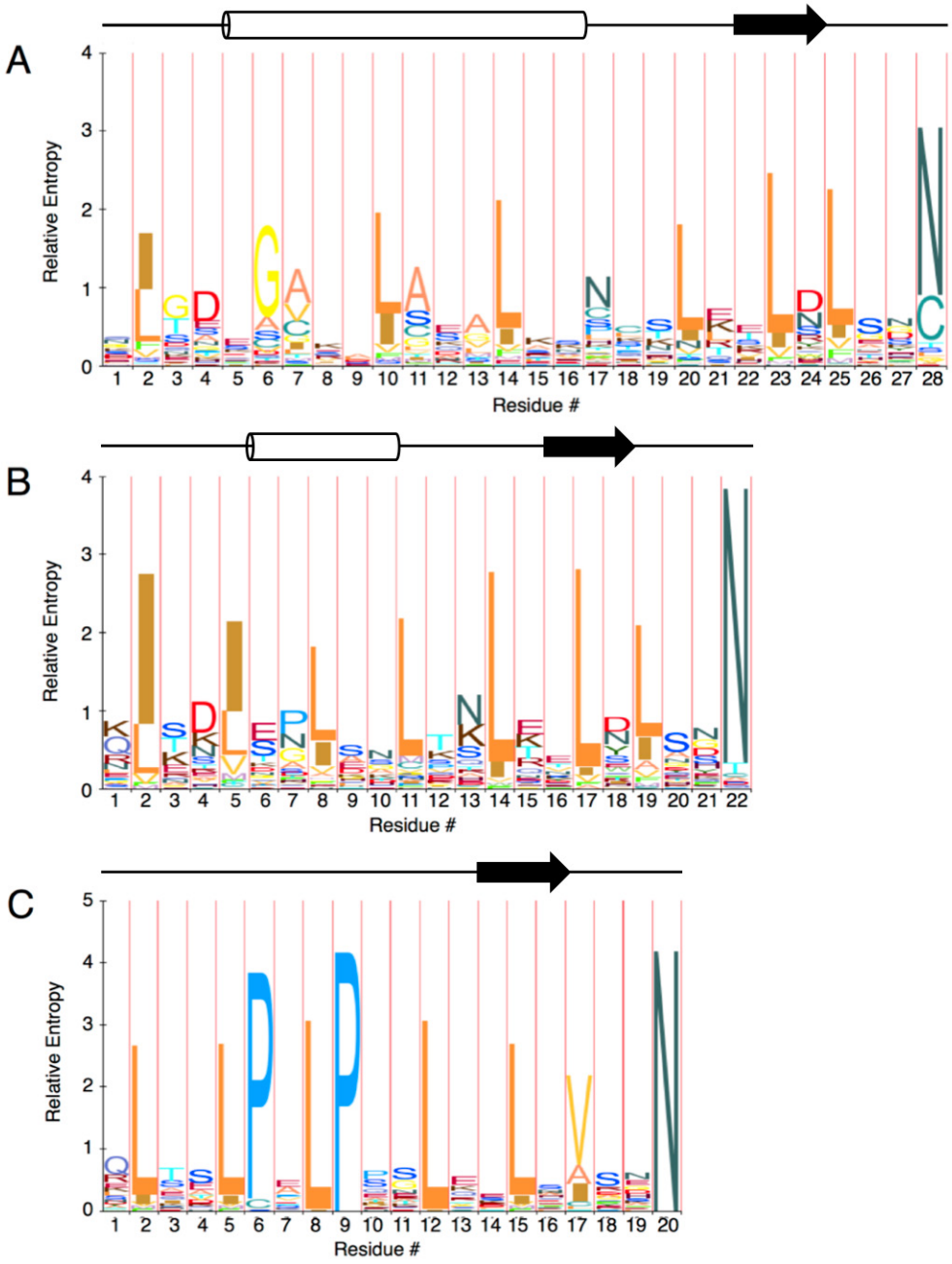


Figure 1.3. Hidden Markov models of representative LRR subfamilies. (A) RI-like subfamily. (B) SD22 subfamily. (C) Bacterial subfamily. The secondary structural elements for each LRR subfamily are shown above each logo.

1.6 Appendix

Urea-induced denaturation of Rna1p

The RI-like LRR protein Rna1p contains 344 amino acids forming 11 repeats (Figure A-1A). Each repeat comprises a β-strand and an α-helix connected by a loop. To determine the stability and the extent of cooperativity within Rna1p, we monitored urea-induced denaturation by CD at 222 nm (Figure A-1B). The denaturation curve is sigmoidal, with a poorly defined native baseline and a broad unfolding transition. Fitting the data using a two-state model yielded an *m*-value of 2.6 kcal • mol^{-1} • M^{-1}. For a protein of the size of Rna1p to unfold by a two-state mechanism, the *m*-value is expected to be 3.8 kcal • mol^{-1} • M^{-1} or higher (Myers et al., 1995). Therefore, Rna1p likely unfolds by a multistate equilibrium mechanism.

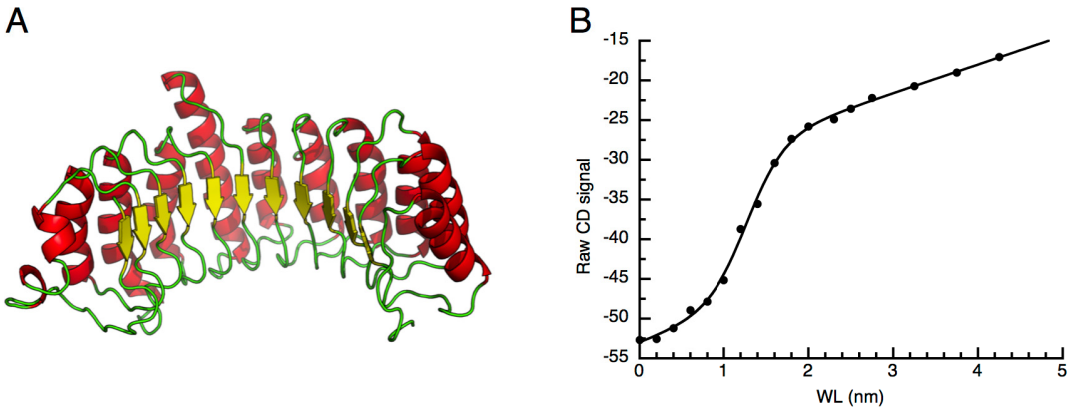


Figure A-1. Structure and urea-induced denaturation of Rna1p. (A) The crystal structure of Rna1p (PDB: 1YRG). α-helices are in red; β-strands are in yellow; coil regions are in green. (B) Urea-induced unfolding transition monitored by CD at 222 nm (filled circles). Line results from fitting a two-state unfolding model to the data.

Chapter II

Capping motifs stabilize the LRR protein PP32 and rigidify adjacent repeats[§]

[§] This chapter is reprinted with permission from the authors. T.P. Dao, A. Majumdar and D. Barrick (2014). *Protein Science* **23**, 801-11.

2.1 Abstract

Capping motifs are found to flank most β-strand-containing repeat proteins. To better understand the roles of these capping motifs in organizing structure and stability, we carried out folding and solution NMR studies on the leucine-rich repeat (LRR) domain of PP32, which is composed of five tandem LRR, capped by α-helical and β–hairpin motifs on the N- and C-termini. We were able to purify PP32 constructs lacking either cap and containing destabilizing substitutions. Removing the C-cap results in complete unfolding of PP32. Removing the N-cap had a much less severe effect, decreasing stability but retaining much of its secondary structure. In contrast, the dynamics and tertiary structure of the first two repeats are significantly perturbed, based on ^{1}H-^{15}N relaxation studies, chemical shift perturbations, and residual dipolar couplings. However, more distal repeats (3 to C-cap) retain their native tertiary structure. In this regard, the N-cap drives the folding of adjacent repeats from what appears to be a molten-globule-like state. This interpretation is supported by extensive analysis using core packing substitutions in the full-length and N-cap-truncated PP32. This work highlights the importance of caps to the stability and structural integrity of β-strand-containing LRR proteins, and emphasizes the different contributions of the N- and C-terminal caps.

2.2 Introduction

Repeat proteins are composed of tandem repeated structural motifs that stack together to form a folded domain. These proteins are stabilized by strong interactions between adjacent repeats, but lack sequence-distant contacts commonly observed in globular proteins (Main et al., 2005a). Despite this absence of sequence-distant contacts, repeat proteins are highly cooperative in their unfolding. This modular architecture and high cooperativity permit dissection of energetic contributions of discrete structural units, making them ideal candidates for folding studies (Kloss et al., 2008).

Among the most common types of repeat motifs are β-strand-containing leucine-rich repeats (LRRs). LRRs have conserved LxxLxLxxN/CxL amino acid sequence motifs (x represents any amino acid), and fold into an extended β-sheet structure such that the conserved hydrophobic leucines point towards the protein core (Kajava, 1998). Most LRR proteins also have capping motifs on either or both termini (Bella et al., 2008). These motifs are hypothesized to serve several purposes. They can shield the hydrophobic core from solvent, preventing aggregation (Richardson and Richardson, 2002). They can become structured in the transition state ensemble, and thus, may guide folding, as was observed for the N-terminal α-helical cap of the LRR domain of InlB (Courtemanche and Barrick, 2008). Terminal caps can also be essential in maintaining structural integrity of the adjacent LRR repeats, as seen for the C-terminal β-strand cap from the 15-repeat LRR protein YopM (Kloss and Barrick, 2009). Finally, the

terminal caps of some LRR domains have been implicated in binding directly to partner proteins (Chen et al., 2008).

One LRR domain that shows recognizable capping structures is the N-terminal domain of PP32, a member of the evolutionarily conserved acidic nuclear phosphoprotein family. The LRR domain of PP32 is composed of five leucine-rich repeats flanked by N-terminal α-helical and C-terminal β–hairpin capping motifs (Figure 2.1). Both caps are highly conserved in primary structure. The N-terminal cap has the same structures in three closely related constructs (de Chiara et al., 2008; Huyton and Wolberger, 2007; Matilla and Radrizzani, 2005; Tochio et al., 2010). The C-terminal cap also shows similar structural features, although the constructs differ in their C-terminal boundaries, complicating detailed structural comparison (Figures 2.1B, C). Alignment of C-terminal caps in a subfamily of LRR proteins including PP32 reveals a consensus sequence **Y**RxxϕxxxϕPxϕxxL**D** (ϕ represents a hydrophobic residue, x represents any residue) (Ceulemans et al., 1999). In PP32, the flanking residues Y and D (Y131 and D146) form a structural hydrogen bond (Figures 2.1B, C). This conserved C-terminal sequence has been proposed to be necessary for nuclear targeting (Stone et al., 1993), and to shield the hydrophobic core from solvent, as well as contribute to the overall stability of the LRR domain.

To test whether N- and C-terminal caps are essential for folding and for maintaining short-range structural integrity in the LRR protein PP32, we have created several PP32 variants with modified capping structures and examined

the effects on equilibrium stability and overall structural integrity. Deleting the C-cap, which removes the acceptor D146 of the conserved hydrogen bond, completely unfolds PP32. However, the disruption of the hydrogen bond alone is not responsible for this unfolding, since construct PP32 Y131F/D146L is folded and has highly similar structure to wild-type PP32. A construct lacking the N-cap (Δ_{NCap} PP32) is destabilized across the LRR domain, but retains significant secondary structure. Solution NMR and core packing mutational studies of Δ_{NCap} and full-length PP32 show that, in addition to providing long-range stability, the N-terminal cap is required for structural integrity of the two adjacent LRRs (one and two).

2.3 Results

Defining the C-terminus of the PP32 LRR domain

Members of the Anp32 family contain a highly conserved N-terminal LRR domain, followed by a variable C-terminal acidic region (Figure 2.1A). The boundaries of the LRR domain of PP32 (human Anp32A) were first suggested by x-ray crystallography of a construct spanning residues 1-149 (Figures 2.1B, 2.1D, Huyton and Wolberger, 2007). In this structure, the C-terminal residues form a single β-strand β7 (residues 144-145). Subsequent NMR studies of two PP32 homologs, mouse Anp32A (de Chiara et al., 2008) and human Anp32B (Tochio et al., 2010), included C-terminal residues up to 164, where the acidic region begins. Some of these additional residues form a well-ordered structural motif that appears to stabilize a C-terminal β-hairpin (residues 144-145 {β-strands β7} pairing with residues 148-149 {β-strands β8}; cyan, Figure 2.1C). NMR ^{15}N relaxation studies showed both proteins to be well structured from residues 1-154 (de Chiara et al., 2008; Tochio et al., 2010). Therefore, the added residues may stabilize the C-terminal capping motif for the LRR region of PP32.

To determine whether these additional C-terminal sequence elements are important for structure and stability, we compared secondary structures and folding properties of constructs 1-145, 1-149 and 1-154 (Figure 2.1D) from human Anp32A (hAnp32A). Far-UV CD spectra for both hAnp32A 1-149, and 1-154 show minima at 217 nm (Figure 2.2A), characteristic of proteins with β-sheet structure. The two spectra also have nearly identical shapes, indicating

similar secondary structures. However, hAnp32A 1-154 displays a stronger negative ellipticity, consistent with ordering of the β-hairpin structure (β7 and β8) in the 154-residue construct, as observed by NMR (de Chiara et al., 2008; Tochio et al., 2010). Surprisingly, the CD spectrum of hAnp32A 1-145 shows little signal at 217 nm but instead a minimum around 200 nm, characteristics of an unfolded protein, suggesting that residues 146-149 (which include the conserved hydrogen bond) are critical for maintaining the folded structure.

To determine the effect of residues 150-154 on stability, urea-induced unfolding transitions for hAnp32A 1-149 and 1-154 were monitored by far-UV CD. Both denaturation curves are sigmoidal and can be well fitted by an equilibrium two-state model in which the unfolding free energy is linearly dependent on urea concentration (Figure 2.2B). The fitted *m*-values for both constructs are the same within error (Table 2.1). As *m*-values have been shown to be correlated with the size of cooperative unit (Myers et al., 1995), hAnp32A 1-149 and 1-154 appear to be undergoing structural transitions of the same size. The magnitudes of the fitted *m*-values (2.86 and 2.88 kcal mol^{-1} M^{-1}) suggest both proteins unfold in a single concerted reaction.[**] The C_m and $\Delta G^o_{H_2O}$ values for hAnp32 1-154 are substantially higher than those for hAnp32A 1-149 (Figure 2.2B, Table 2.1), indicating that the C-terminal residues 150-154 contribute significantly to the stability of the LRR domain of hAnp32A (by more than 2 kcal • mol^{-1}). C_m and

[**] From a comparative study of two-state unfolding of globular proteins, the *m*-value for a protein the size of hAnp32A is predicted to be 1.84 kcal mol^{-1} M^{-1} or higher (Myers et al., 1995).

$\Delta G^{o}_{H_2O}$ values for hAnp32A 1-161 are the same as for 1-154 (data not shown), consistent with observation from the NMR studies that residues following 154 are highly dynamic and unstructured (de Chiara et al., 2008; Tochio et al., 2010). These observations provide a thermodynamic boundary for the LRR domain of the hAnp32A (residues 1-154). For the remainder of this work, we will consider construct 1-154 as the hAnp32A LRR domain, which we will refer to as "PP32" for brevity.

The role of the conserved hydrogen bond between Y131 and D146

Removing only four C-terminal residues from construct hAnp32A 1-149 leads to global unfolding. The resulting C-terminal deletion construct, hAnp32A 1-145, cannot form the conserved hydrogen bond in the C-terminal cap since it is missing the acceptor D146. To determine if this hydrogen bond is essential to the folding and stability of PP32, we replaced Y131 with phenylalanine and D146 with leucine (constructs Y131F, D146L, and Y131F/D146L) in the hAnp32A 1-154 construct. Unlike the hAnp32 1-145 construct, these variants are folded by far-UV CD (Figure 2.3A). To determine if there are significant changes in tertiary structure upon removal of the conserved hydrogen bond, we compared ^{15}N-^{1}H-HSQC spectra of wild-type PP32 and Y131F/D146L. The HSQC spectra reveal well-dispersed resonances, indicating that both proteins are well-folded (Figure 2.3B). Overall, the spectrum of Y131F/D146L looks similar to that of wild-type PP32, with many of the peaks overlaying. However, some peaks are perturbed in the double variant.

To map these chemical shift changes, we assigned the resonances for wild-type PP32 using standard 3D NMR experiments and for Y131F/D146L by comparison to wild-type PP32. The chemical shift perturbations (CSPs) are mostly small and localized to the sites of substitution (residues 131 and 146, Figure 2.3C). The N-terminal cap and first four LRRs show little to no chemical shift changes. The high chemical shift dispersion across the entire domain, along with the restriction of chemical shift perturbations to the site of the substitutions, suggests that the overall LRR fold is maintained, despite the removal of this conserved hydrogen bond.

To determine the energetic consequences of removing the conserved hydrogen bond on folding, urea-induced unfolding transitions for the variants were monitored by far-UV CD (Figure 2.3D). Based on the decreased $\Delta G^{o}_{H_2O}$ and C_m values, these substitutions are all destabilizing, indicating that the folded state of PP32 is indeed stabilized by this capping hydrogen bond (Table 2.1). The differences in the extent of destabilization for Y131F and D146L single site variants ($\Delta\Delta G^{o}_{H_2O}$ of 1 kcal•mol^{-1} and 3.5 kcal•mol^{-1}, respectively) indicate additional interactions between these two residues and their intermediate surroundings. Moreover, the free energy changes between individual substitutions are not nonadditive, with the larger destabilization in the wild-type background, compared to the singly substituted backgrounds. This nonadditivity suggests that the true contribution of the hydrogen bond may be overestimated from analysis of the single mutants, which likely pay a penalty for desolvation and

burial of a hydrogen bonding group. All three variants have nearly identical *m*-values to wild-type PP32 (Figure 2.3D, Table 2.1), suggesting fully cooperative unfolding transitions, despite the highly destabilizing effects.

The role of the conserved N-terminal α-helical capping motif

To investigate the structural and energetic contributions of the N-terminal capping motif, we removed the 18 residues prior to the first repeat (Δ_{Ncap} PP32). Far-UV CD spectra for full-length and Δ_{Ncap} PP32 have nearly identical shapes, suggesting that the two proteins have similar secondary structures (data not shown). Urea-induced unfolding reveals that Δ_{NCap} PP32 has lower $\Delta G^{o}_{H_2O}$ and C_m values than full-length PP32 (Figure 2.4A, Table 2.1), indicating that PP32 is less stable without the N-terminal capping motif. Unlike the deletion of residues 150-154, the *m*-value for Δ_{NCap} PP32 decreases, consistent with a smaller cooperative unfolding unit.

To obtain a more detailed structural picture of the consequences of N-terminal cap deletion, we studied Δ_{Ncap} PP32 by NMR spectroscopy. The HSQC spectrum shows a large number of well-dispersed resonances, indicative of a folded protein (Figure 2.4B). However, comparison of full-length and Δ_{NCap} PP32 HSQC spectra shows that many peaks have significantly shifted. To map these chemical shift changes, we assigned the resonances for Δ_{NCap} PP32. Large CSPs were observed for N-terminal repeats one and two, whereas the last three repeats and C-terminal capping motif remain largely unperturbed (Figures

2.4C, 2.4D). Using the program TALOS+ (Cornilescu et al., 1999), which predicts secondary structures based on chemical shifts and sequence information, we observe highly similar secondary structure distributions in full-length and Δ_{NCap} PP32 despite significant chemical shift changes in repeats one and two. In particular, the β-strands of the LRRs are retained along the entire Δ_{NCap} PP32 construct.

To determine whether the large CSPs simply resulted from a local change in chemical environment upon removal of the N-terminal capping motif, or from a larger conformational rearrangement of repeats one and two, we measured residual dipolar couplings (RDCs) for both PP32 (data not shown) and Δ_{NCap} PP32 (Figure 2.5A). RDCs provide information on orientations between directly bonded atoms and the overall molecular frame, and thus, present a sensitive means to compare the higher-order structural details of related proteins. From repeat two to the C-terminus, the strong correlation between the RDCs of full-length and Δ_{NCap} PP32 indicates high structural similarity (Figure 2.5B). In contrast, for the first repeat, there is relatively poor correlation between the RDCs of full-length and Δ_{NCap} PP32, suggesting that the structure of the first repeat has undergone a significant conformational rearrangement.

To directly probe whether the conformational changes increase dynamics upon removing the N-terminal cap, we carried out relaxation studies on the Δ_{NCap} PP32 (Figure 2.6). R_1 values are nearly constant along the sequence. However, R_2 values are elevated for the first 55 residues, but converge to a constant value

towards the C-terminus. These data are consistent with μs-ms-timescale fluctuation in the N-terminal two repeats. Similar to the R_2 profile, heteronuclear NOE intensities are high and uniform (around the theoretical value of 0.8) for most of the protein, except for the first 35 residues on the N-terminus, where the intensities are varied and slightly depressed. These observations indicate that the C-terminal region of Δ_{NCap} PP32 is folded and rigid whereas the first repeat and half of the second repeat experience fast backbone fluctuation on the ps-ns timescale. For mouse Anp32A (de Chiara et al., 2008) and human Anp32B (Tochio et al., 2010), the R_1, R_2 and heteronuclear NOE profiles have been shown to be relatively constant along the entire domain (including the N-terminal α-helical capping motif and the first LRR). Consistent with these changes in dynamics, we observed an increase in the intensities of many N-terminal HSQC resonances of Δ_{NCap} PP32 with increasing temperature (data not shown). The enhanced R_2 and depressed heteronuclear NOE values seen here indicate that the α-helical cap rigidifies LRRs one and two.

Removal of the N-terminal α-helical capping motif disrupts side chain packing of the N-terminal LRRs

To determine the extent to which deletion of the N-terminal cap disrupts side-chain packing, we made conservative single-site Leu→Ala and Val→Gly substitutions at core positions of both full-length and Δ_{Ncap} PP32. Far-UV CD spectra of full-length PP32 packing variants are nearly identical in shape to those

of wild-type PP32, indicating that the β-sheet structure of PP32 is retained (Figure 2.7A). However, spectra of many of the same variants of Δ_{NCap} PP32 differ from that of the parent construct. These differences depend on the location of the substitutions. Packing substitutions in the C-terminus shift spectra towards random coil, whereas substitutions in the N-terminus show less pronounced spectral broadening (Figure 2.7A).

To quantitatively assess the effects of these packing substitutions on structural stability, urea-induced unfolding transitions were monitored by far-UV CD (Figure 2.7B). Based on decreases in C_m, full-length PP32 variants are significantly destabilized across the entire domain (Table 2.2). C-terminal substitutions (L83A, L109A and V135G) are similarly destabilized in the Δ_{NCap} background.[††] However, substitutions in the first two repeats of Δ_{NCap} (L22A, L37A, L47A and L60A) produce variants with about the same stabilities as the parent (Δ_{NCap}) construct. These four sites are in the region of increased dynamics (based on R_2 and heteronuclear NOE) and structural change (based on CSPs and RDCs). These data are consistent with a structural ensemble in which packing is disrupted in the first two LRRs of Δ_{NCap} PP32, but retained in LRR3 through the C-terminus.

[††] No native baselines for these constructs were observed and, therefore, their unfolding free energies cannot be quantified.

2.4 Discussion

Several studies have examined the contributions of individual units to the structure and stability of linear repeat proteins through addition and removal of repeats (Aksel et al., 2011; Kloss and Barrick, 2009; Kloss et al., 2008; Main et al., 2005b; Mello and Barrick, 2004; Tripp and Barrick, 2007; Vieux and Barrick, 2011; Wetzel et al., 2008). Most studies have focused on α-helical repeat proteins, which appear to be tolerant of manipulations of whole repeats (Mello and Barrick, 2004; Sue et al., 2008; Tripp and Barrick, 2007). In contrast, deletion of β-strand-containing repeat motifs could result in large-scale disruption of adjacent (and non-adjacent) repeats. Removing the C-terminal motifs from YopM caused unfolding of several preceding repeats (Kloss and Barrick, 2009). The prevalence of terminal capping structures in β-strand-containing repeat proteins (and LRR proteins in particular) may serve as a means to protect otherwise labile β-strand repeats. Removing the N-terminal caps from β-strand-containing LRR proteins InlB and YopM resulted in disordered polypeptides (N. Courtemanche, E. Kloss and DB, unpublished). Deletion of the C-terminal cap from pertactin β-helix resulted in mixture of aggregates and soluble oligomers (Bryan et al., 2011). In this study, we seek to further understand the roles of capping motifs on structures and stabilities of LRR proteins.

C-terminal capping structures are essential for stability and the structural integrity of the entire PP32 LRR domain

Despite making little contact with the rest of the protein, the β-hairpin cap on the C-terminus of PP32 is critical for stability and folding (Figure 2.2). Removing the last five residues (the second strand of the β-hairpin, hAnp32A 1-149) decreases the stability of PP32 by more than 2 kcal/mol (Table 2.1). The destabilizing effect of removing the next four residues (the first strand of the β-hairpin, hAnp32A 1-145) is greater than the overall stability, thus resulting in global unfolding even under native conditions. This is a more global structural disruption than that produced by a similar C-terminal deletion in YopM, where removing the β-cap causes only the three preceding LRRs to unfold (Kloss and Barrick, 2009). This complete loss of structure explains the loss of function in yeast SDS22, where deletion of the C-terminal cap disrupts nuclear localization (Stone et al., 1993). It is important to note that the increased tolerance of YopM to β-cap removal might be a result of the larger size of YopM (15 repeats). Unlike the pertactin β-helix, PP32 remains monomeric upon C-terminal cap deletion (Bryan et al., 2011). For PP32, the cap protects the structural integrity of the entire molecule. Global unfolding in hAnp32A 1-145 is not solely the result of loss of the conserved hydrogen bond between Y131 and D146 (Figure 2.3), highlighting the importance of other interactions within the C-terminal cap.

The N-terminal cap stabilizes and constrains the first two repeats

In contrast to LRR proteins InlB and YopM, PP32 is folded and well-behaved without its N-terminal cap. In addition to being less stable (Figure 2.4A), Δ_{Ncap} PP32 has lower than predicted *m*-value based on the number of residues

removed[‡‡], indicating that the cooperative unit for Δ_{Ncap} PP32 is smaller than the construct itself. However, Δ_{Ncap} PP32 has similar secondary structure to the full-length PP32, based on far-UV CD and TALOS (Cornilescu et al., 1999, data not shown). Interestingly, the RDC values of the two constructs correlate well, except for the first repeat, suggesting that the first repeat of Δ_{NCap} PP32, as a unit, might align differently than in full-length PP32, relative to the rest of the LRR domain (Figure 2.5B). Relaxation studies show that the first two LRRs experience increased dynamics both in fast and slow timescales upon N-terminal cap deletion (Figure 2.6). Together, the data point to a less ordered, or packed, N-terminus, but well-folded native-like C-terminus in Δ_{NCap} PP32, as confirmed by mutational analysis of the two constructs (Figure 2.7). Upon removal of N-terminal cap, PP32 seems to have molten globule-like structure, a partially folded intermediate where secondary structures are formed but rigid tertiary packing is disrupted (Ohgushi and Wada, 1983; Ptitsyn et al., 1990).

A possible mechanism for the folding of PP32

We have been able to obtain folded PP32 that is missing a significant segment of its N-terminus (18 residues, Δ_{Ncap} PP32, and up to 65 residues [data not shown]). However, removing only four C-terminal residues (146-149) significantly destabilizes the protein, suggesting that the C-terminus might be the most stable region of the molecule. Studies of many globular proteins, and

[‡‡] Removing the N-terminal capping motif of PP32 reduces its length by 12%, but its *m*-value by 18%.

α-helical repeat proteins (myotrophin and Notch ankyrin) have shown that folding pathways are determined by local stabilities (Bradley and Barrick, 2006; Hoang et al., 2002; Jennings and Wright, 1993; Lowe and Itzhaki, 2007; McCallister et al., 2000; Nauli et al., 2001; Raschke and Marqusee, 1997; Tripp and Barrick, 2007). Based on these observations, our results suggest that for the LRR domain of PP32, the C-terminus folds first, and may nucleate the rest of the fold. This proposed mechanism differs from the LRR domain of InlB, which folds through a polarized N-terminal pathway involving the N-cap and first two repeats. The C-terminal folding mechanism for PP32 seems at odds with a cotranslational folding model, as folding of the nascent PP32 chain will be prevented until the C-terminus is released from the exit tunnel.

2.5 Materials and Methods

Subcloning, protein expression, and purification

The gene encoding the PP32 was a kind gift from the laboratory of Dr. Cynthia Wolberger. LRR encoding segments were subcloned into the NdeI and XhoI sites of pET24b (Novagen, Madison, WI) using PCR. A tryptophan codon was added to facilitate protein concentration determination. Point substitutions were made using Quikchange (Stratagene, La Jolla, CA).

Constructs were expressed in *Escherichia coli* BL21 cells in autoinduction media (Studier, 2005) at 20^{o}C overnight. Bacteria were pelleted, lysed in 20 mM sodium phosphate, 500 mM NaCl, 25 mM imidazole, 0.1 mM TCEP (pH 7.4), and purified via Ni^{2+} chromatography. Purified proteins were dialyzed into 20mM sodium phosphate, 150 mM NaCl, 0.1 mM TCEP (pH 7.8), and were frozen at -80^{o}C. Protein concentrations were determined as described (Edelhoch, 1967).

Circular Dichroism spectroscopy

All CD measurements were carried out using Aviv Model 400 CD spectrometer (Lakewood, NJ). Far-UV CD spectra were collected in an 0.1 cm path-length quartz cuvette with protein concentrations ranging from 15 to 30 μM. Spectra were averages of three wavelength scans, each with 1 nm step size and 10 seconds signal average per step.

Equilibrium unfolding

Urea-induced unfolding was monitored by CD at 220 nm. Urea (Amresco, Solon, OH) was deionized by chromatography over mixed-bed resin (Bio-Rad, Hercules, CA). Urea concentration was determined by refractometry (Pace, Nick C, 1986) Urea titrations were carried out using a computer-controlled Microlab syringe titrator (Hamilton, Reno, NV). Samples contained 2-4 μM protein, 20 mM sodium phosphate, 150 mM NaCl, 0.1 mM TCEP (pH 7.8). At each urea concentration, samples were equilibrated for 5-10 minutes at 20^{o}C and CD signal was averaged for 30 seconds. Two-state analysis of equilibrium unfolding transitions were carried out as described by Street et al (2008).

NMR spectroscopy

^{15}N- and ^{15}N,^{13}C-labeled PP32 and variants were expressed in M9 minimal medium supplemented with $^{15}NH_4Cl$ and ^{13}C-labeled glucose (Cambridge Isotope Laboratories, Andover, MA) at 20^{o}C overnight and purified as described above. NMR samples contained 0.6-1.2 mM proteins, 20 mM sodium phosphate, 50 mM NaCl, 0.1 mM TCEP, 0.1 mM EDTA and 5% D_2O (pH 6.8). For RDC experiments, 0.6 mM full-length and Δ_{NCap} ^{15}N-labelled PP32 constructs were aligned using a liquid crystalline medium containing 5% (w/w) pentaethylene glycol monododecyl ether (C12E5), with 1-hexanol (Sigma-Aldrich, St. Louis, MO) at mole ratio (to C12E5) of r = 0.85 (Rückert and Otting, 2000). NMR experiments comparing PP32 and PP32 Δ_{NCap} were performed at 30^{o}C and those comparing PP32 and PP32 Y131F/D146L were performed at 20^{o}C. Spectra were collected on a Bruker Avance II 600 MHz spectrometer equipped with a

cryoprobe, processed using NMRPipe (Delaglio et al., 1995), and displayed and analyzed with Sparky (Goddard, T. D. and Kneller, D. G.).

Resonance assignments for PP32 and PP32 Δ_{NCap} were carried out using triple-resonance spectra, including HNCACB, CBCA(CO)NH, HNN, HBHA(CO)NH, HNCA, and ^{15}N-edited HSQC- NOESY. Assignments were determined using CARA.(Masse and Keller, 2005) Chemical shift perturbations (CSPs) between constructs were determined from differences in resonance positions in ^{1}H–^{15}N HSQC spectra using the equation $\Delta\delta = [(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2]^{1/2}$, where $\Delta\delta_H$ and $\Delta\delta_N$ are differences in the ^{1}H and the ^{15}N chemical shifts, respectively, for a given residue. RDC values for full-length and Δ_{NCap} PP32 were each measured with two sets of IPAP experiments in alignment media and in isotropic media (Ottiger et al., 1998). Peak positions in the IPAP-HSQC spectra were determined using the program PATI (Berlin et al., 2009). R_1, R_2, and heteronuclear NOE measurements for Δ_{NCap} PP32 were collected using standard pulse sequences. ^{15}N T_1 experiments used relaxation delays of 97.5, 195, 299 (duplicated), 494, 696, 897 and 1099 ms. ^{15}N T_2 experiments used relaxation delays of 7.6, 15.2, 22.8, 30.4 (duplicated), 38, 45.6, and 53.2 ms. Relaxation rates were determined using the program RELAXFIT, with 500 Monte Carlo trials for error analyses (Fushman et al., 1997).

2.6 References

Aksel, T., Majumdar, A., and Barrick, D. (2011). The Contribution of Entropy, Enthalpy, and Hydrophobic Desolvation to Cooperativity in Repeat-Protein Folding. Structure *19*, 349–360.

Bella, J., Hindle, K.L., McEwan, P.A., and Lovell, S.C. (2008). The leucine-rich repeat structure. Cell. Mol. Life Sci. *65*, 2307–2333.

Berlin, K., O'Leary, D.P., and Fushman, D. (2009). Improvement and Analysis of Computational Methods for Prediction of Residual Dipolar Couplings. J Magn Reson *201*, 25–33.

Bradley, C.M., and Barrick, D. (2006). The Notch Ankyrin Domain Folds via a Discrete, Centralized Pathway. Structure *14*, 1303–1312.

Bryan, A.W., Starner-Kreinbrink, J.L., Hosur, R., Clark, P.L., and Berger, B. (2011). Structure-based prediction reveals capping motifs that inhibit β-helix aggregation. PNAS *108*, 11099–11104.

Ceulemans, H., De Maeyer, M., Stalmans, W., and Bollen, M. (1999). A capping domain for LRR protein interaction modules. FEBS Letters *456*, 349–351.

Chen, S., Li, B., Grundke-Iqbal, I., and Iqbal, K. (2008). I PP2A 1 Affects Tau Phosphorylation via Association with the Catalytic Subunit of Protein Phosphatase 2A. J Biol Chem *283*, 10513–10521.

De Chiara, C., Menon, R.P., and Pastore, A. (2008). Structural bases for recognition of Anp32/LANP proteins. FEBS Journal *275*, 2548–2560.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR *13*, 289–302.

Courtemanche, N., and Barrick, D. (2008). The leucine-rich repeat domain of Internalin B folds along a polarized N-terminal pathway. Structure *16*, 705–714.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry *6*, 1948–1954.

Fushman, D., Cahill, S., and Cowburn, D. (1997). The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of 15N

relaxation with monomer/dimer equilibration. Journal of Molecular Biology *266*, 173–194.

Goddard, T. D., and Kneller, D. G. SPARKY J.

Hoang, L., Bédard, S., Krishna, M.M.G., Lin, Y., and Englander, S.W. (2002). Cytochrome c folding pathway: Kinetic native-state hydrogen exchange. PNAS *99*, 12173–12178.

Huyton, T., and Wolberger, C. (2007). The crystal structure of the tumor suppressor protein pp32 (Anp32a): Structural insights into Anp32 family of proteins. Protein Science *16*, 1308–1315.

Jennings, P.A., and Wright, P.E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science *262*, 892–896.

Kajava, A.V. (1998). Structural diversity of leucine-rich repeat proteins. Journal of Molecular Biology *277*, 519–527.

Kloss, E., and Barrick, D. (2009). C-terminal deletion of leucine-rich repeats from YopM reveals a heterogeneous distribution of stability in a cooperatively folded protein. Protein Sci *18*, 1948–1960.

Kloss, E., Courtemanche, N., and Barrick, D. (2008). Repeat-protein folding: New insights into origins of cooperativity, stability, and topology. Archives of Biochemistry and Biophysics *469*, 83–99.

Lowe, A.R., and Itzhaki, L.S. (2007). Rational redesign of the folding pathway of a modular protein. PNAS *104*, 2679–2684.

Main, E.R., Lowe, A.R., Mochrie, S.G., Jackson, S.E., and Regan, L. (2005a). A recurring theme in protein engineering: the design, stability and folding of repeat proteins. Current Opinion in Structural Biology *15*, 464–471.

Main, E.R.G., Stott, K., Jackson, S.E., and Regan, L. (2005b). Local and long-range stability in tandemly arrayed tetratricopeptide repeats. PNAS *102*, 5721–5726.

Masse, J.E., and Keller, R. (2005). AutoLink: Automated sequential resonance assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. Journal of Magnetic Resonance *174*, 133–151.

Matilla, A., and Radrizzani, M. (2005). The Anp32 family of proteins containing leucine-rich repeats. Cerebellum *4*, 7–18.

McCallister, E.L., Alm, E., and Baker, D. (2000). Critical role of β-hairpin formation in protein G folding. Nat Struct Mol Biol *7*, 669–673.

Mello, C.C., and Barrick, D. (2004). An experimentally determined protein folding energy landscape. PNAS *101*, 14102–14107.

Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci *4*, 2138–2148.

Nauli, S., Kuhlman, B., and Baker, D. (2001). Computer-based redesign of a protein folding pathway. Nat Struct Mol Biol *8*, 602–605.

Ohgushi, M., and Wada, A. (1983). "Molten-globule state": a compact form of globular proteins with mobile side-chains. FEBS Letters *164*, 21–24.

Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement ofJand Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. Journal of Magnetic Resonance *131*, 373–378.

Pace, Nick C (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol. *131*, 266–280.

Ptitsyn, O.B., Pain, R.H., Semisotnov, G.V., Zerovnik, E., and Razgulyaev, O.I. (1990). Evidence for a molten globule state as a general intermediate in protein folding. FEBS Letters *262*, 20–24.

Raschke, T.M., and Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. Nat Struct Mol Biol *4*, 298–304.

Richardson, J.S., and Richardson, D.C. (2002). Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. PNAS *99*, 2754–2759.

Rückert, M., and Otting, G. (2000). Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. J. Am. Chem. Soc. *122*, 7793–7797.

Stone, E.M., Yamano, H., Kinoshita, N., and Yanagida, M. (1993). Mitotic regulation of protein phosphatases by the fission yeast sds22 protein. Current Biology *3*, 13–26.

Street, T.O., Courtemanche, N., and Barrick, D. (2008). Protein Folding and Stability Using Denaturants. In Methods in Cell Biology, I. Dr. John J. Correia and Dr. H. William Detrich, ed. (Academic Press), pp. 295–325.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Sue, S.-C., Cervantes, C., Komives, E.A., and Dyson, H.J. (2008). Transfer of flexibility between ankyrin repeats in IkappaB* upon formation of the NF-kappaB complex. J. Mol. Biol. *380*, 917–931.

Tochio, N., Umehara, T., Munemasa, Y., Suzuki, T., Sato, S., Tsuda, K., Koshiba, S., Kigawa, T., Nagai, R., and Yokoyama, S. (2010). Solution Structure of Histone Chaperone ANP32B: Interaction with Core Histones H3–H4 through Its Acidic Concave Domain. Journal of Molecular Biology *401*, 97–114.

Tripp, K.W., and Barrick, D. (2007). Enhancing the Stability and Folding Rate of a Repeat Protein through the Addition of Consensus Repeats. Journal of Molecular Biology *365*, 1187–1200.

Vieux, E.F., and Barrick, D. (2011). Deletion of internal structured repeats increases the stability of a leucine-rich repeat protein, YopM. Biophysical Chemistry *159*, 152–161.

Wetzel, S.K., Settanni, G., Kenig, M., Binz, H.K., and Plückthun, A. (2008). Folding and Unfolding Mechanism of Highly Stable Full-Consensus Ankyrin Repeat Proteins. Journal of Molecular Biology *376*, 241–257.

Table 2.1. Thermodynamic unfolding parameters for hAnp32 1-149, hAnp32A 1-154 and cap variants

	$\Delta G^{o}_{H_2O}$	*m-value*	C_M	$\Delta\Delta G^{o}_{H_2O}$
hAnp32A 1-149	-5.72 ± 0.05	2.88 ± 0.09	1.99 ± 0.05	2.21
hAnp32A 1-154 (PP32)	-7.93 ± 0.18	2.86 ± 0.02	2.77 ± 0.04	NA
PP32 Y131F	-6.97 ± 0.09	3.05 ± 0.04	2.28 ± 0.01	0.96
PP32 D146L	-3.49 ± 0.26	3.05 ± 0.22	1.15 ± 0.02	4.44
PP32 Y131F/D146L	-4.72 ± 0.14	2.79 ± 0.14	1.69 ± 0.06	3.21
Δ_{NCap} PP32	-4.13 ± 0.09	2.37 ± 0.04	1.74 ± 0.02	3.84

Values determined from urea-induced denaturation. $\Delta G^{o}_{H_2O}$ in kcal • mol^{-1}; *m* values in kcal • mol^{-1} • M^{-1}, C_M in M (urea).

Table 2.2. Thermodynamic unfolding parameters for core packing substitutes in full-length and Δ_{NCap} PP32

	$\Delta G^{o}_{H_2O}$	*m-value*	C_M	$\Delta\Delta G^{o}_{H_2O}$
PP32	-7.93 ± 0.18	2.86 ± 0.02	2.77 ± 0.04	NA
PP32 L22A	-3.88 ± 0.05	2.36 ± 0.03	1.64 ± 0.01	4.05
PP32 L37A	-3.68 ± 0.15	2.28 ± 0.07	1.62 ± 0.01	4.25
PP32 L47A	-2.98 ± 0.17	2.22 ± 0.06	1.34 ± 0.05	4.95
PP32 L60A	-3.98 ± 0.10	2.62 ± 0.01	1.52 ± 0.04	3.95
PP32 L83A	-4.46 ± 0.02	3.25 ± 0.02	1.37 ± 0.01	3.47
PP32 L109A	-4.56 ± 0.08	3.37 ± 0.17	1.35 ± 0.06	3.37
PP32 V135G	-4.07 ± 0.19	3.61 ± 0.16	1.13 ± 0.01	3.86
Δ_{NCap} PP32	-4.13 ± 0.09	2.37 ± 0.04	1.74 ± 0.02	3.80
Δ_{NCap} PP32 L22A	-4.07 ± 0.25	2.34 ± 0.11	1.64 ± 0.03	3.86
Δ_{NCap} PP32 L37A	-3.58 ± 0.16	2.23 ± 0.08	1.60 ± 0.03	4.35
Δ_{NCap} PP32 L47A	-2.76 ± 0.12	2.01 ± 0.07	1.37 ± 0.04	5.17
Δ_{NCap} PP32 L60A	-3.03 ± 0.37	2.14 ± 0.28	1.42 ± 0.04	4.90

Values determined from urea-induced denaturation. $\Delta G^{o}_{H_2O}$ in kcal • mol^{-1}; *m* values in kcal • mol^{-1} • M^{-1}, C_M in M (urea). Parameters for Δ_{NCap} PP32 L83A, L109A and V135G could not be determined.

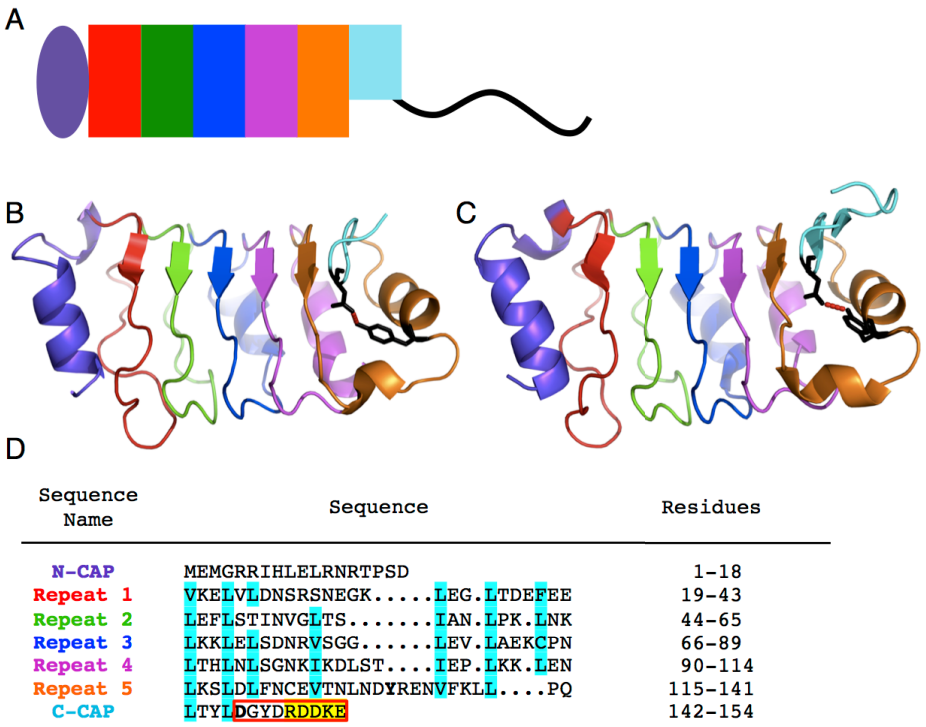


Figure 2.1. Structure and sequence of PP32. (A) Schematic representation of the hAnp32A. The folded N-terminal LRR domain is shown as colored shapes (N-terminal capping motif in purple, repeats 1-5 in red, green, blue, magenta and orange, respectively, the C-terminal capping motif in cyan, and the C-terminal acidic region is shown as a black line. Ribbon representation of (B) the crystal structure of the LRR domain of human Anp32A (PP32) (Huyton and Wolberger, 2007) and (C) the NMR structure of the LRR domain of mouse Anp32A (de Chiara et al., 2008). The conserved hydrogen-bonded Y131 and D146 are shown in black. (D) Sequence alignment of the five LRRs of hAnp32 1-154. Conserved hydrophobic residues and asparagines are highlighted in cyan. The conserved hydrogen bond donor and acceptor Y131 and D146 are bolded. The five C-terminal residues included in this study (residues 150-154) are highlighted in yellow. Construct hAnp32A 1-149 is missing the residues highlighted in yellow. Construct hAnp32A 1-145 is missing residues in the red box.

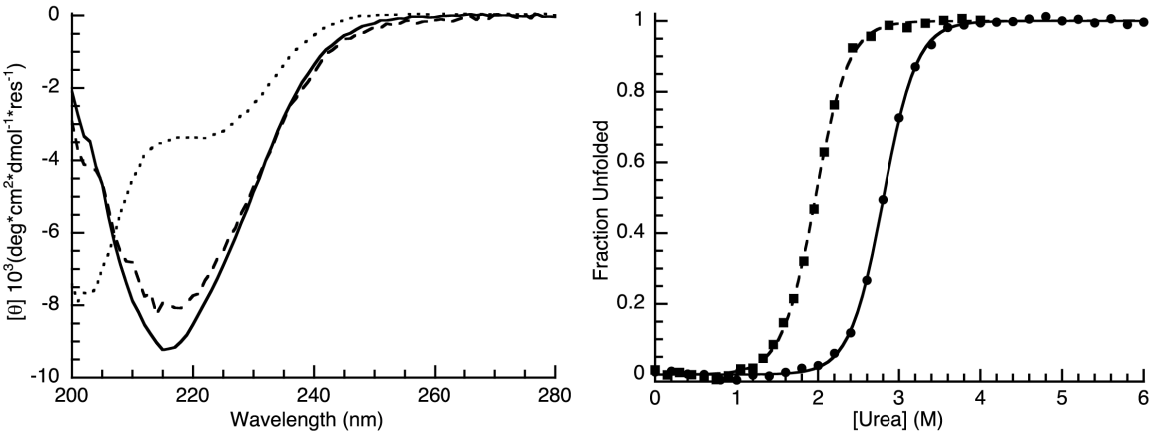


Figure 2.2. CD spectra and equilibrium unfolding of hAnp32A with different C-termini. (A) Far-UV CD shows characteristic β-strand signal with a minimum at 217 nm for both hAnp32A 1-149 (dashed line) and 1-154 (solid line) but disordered polypeptide signal for hAnp32A 1-145 (dotted line). (B) Normalized urea-induced unfolding transitions monitored by far-UV CD for hAnp32A 1-149 (squares, dashed line) and 1-154 (circles, solid line). Lines result from fitting a two-state unfolding model to the data.

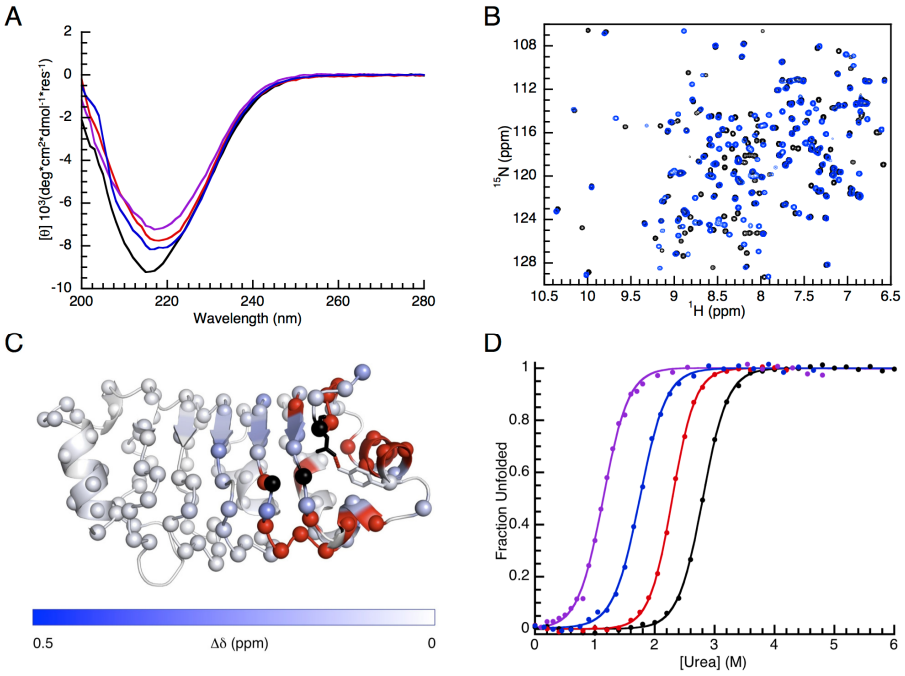


Figure 2.3. Solution spectroscopy and equilibrium unfolding of PP32 and PP32 Y131F/D146L. (A) Far-UV CD shows characteristic β-strand signal with a minimum 217 nm for PP32 (black), and variants Y131F (red), D146L (magenta) and Y131F/D146L (blue). (B) ^{1}H-^{15}N-HSQC spectra of PP32 (black) and PP32 Y131F/D146L (blue) in 20 mM sodium phosphate, 50 mM NaCl, 0.2 mM TCEP, pH 6.8, recorded at 600 MHz and 20°C. (C) Chemical shift perturbations (CSPs) of the PP32 Y131F/D146L variant compared to wild-type as detected in ^{1}H-^{15}N HSQC. Y131 and D146 are represented in sticks. Spheres represent Cα's. CSPs of amide HN_3 for which assignments can be transferred from PP32 to PP32 Y131F/D146L are displayed on a blue to white scale. Residues for which the amide peak disappears are displayed in black, and those that move in a crowded region of the HSQC and therefore cannot be assigned with certainty are displayed in red. (D) Normalized urea-induced unfolding transitions monitored by far-UV CD for PP32, Y131F, D146L, and Y131F/D146L (colors as noted above). Lines result from fitting a two-state unfolding model to the data.

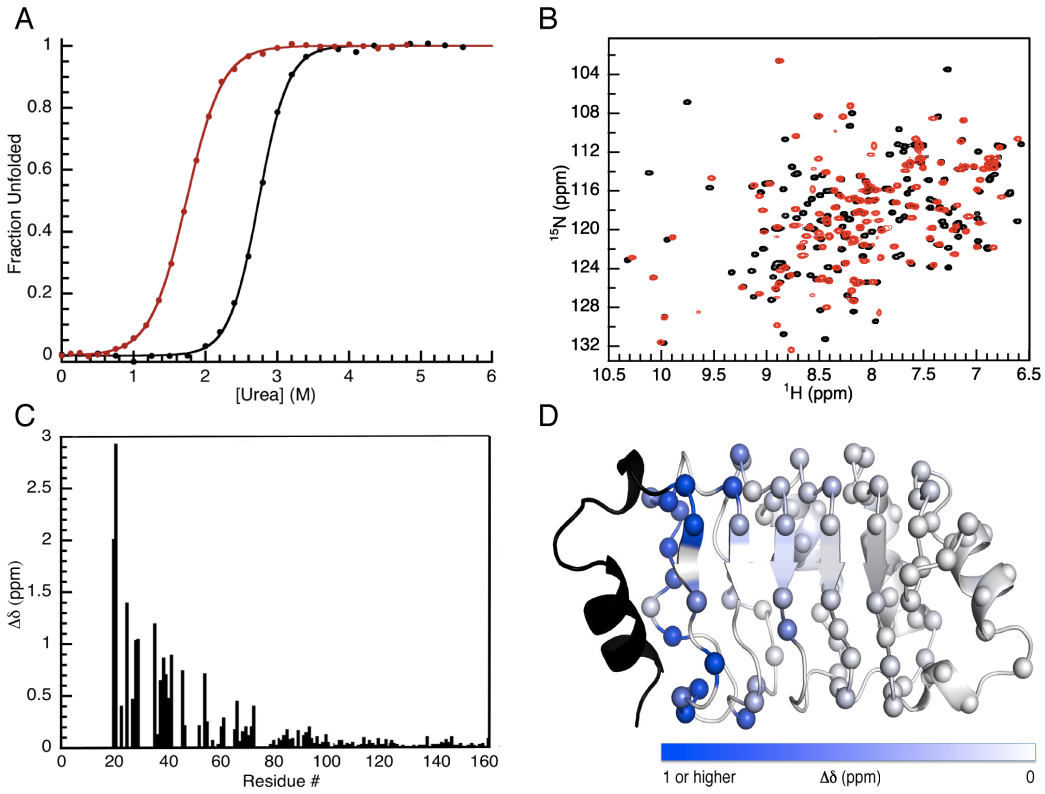


Figure 2.4. Equilibrium unfolding and NMR spectroscopy of PP32 and Δ_{NCap} PP32 (A) Normalized urea-induced unfolding transitions monitored by far-UV CD for PP32 (black) and Δ_{NCap} PP32 (red). Lines result from fitting a two-state unfolding model to the data. (B) Superposition of the ^{1}H-^{15}N HSQC spectra of PP32 (black) and Δ_{NCap} PP32 (red) in 20mM sodium phosphate, 50 mM NaCl, 0.2 mM TCEP, pH 6.8, recorded at 600 MHz and 30°C. (C) CSPs resulting from removal of the N-terminal cap. Residue numbering is based on the PP32 construct. (D) Mapping of CSPs displayed in (C) onto the structure of PP32. The 18-residue α-helical-capping motif is in black. Spheres represent the Cα's of the assigned residues. CSPs are displayed on a white to blue scale.

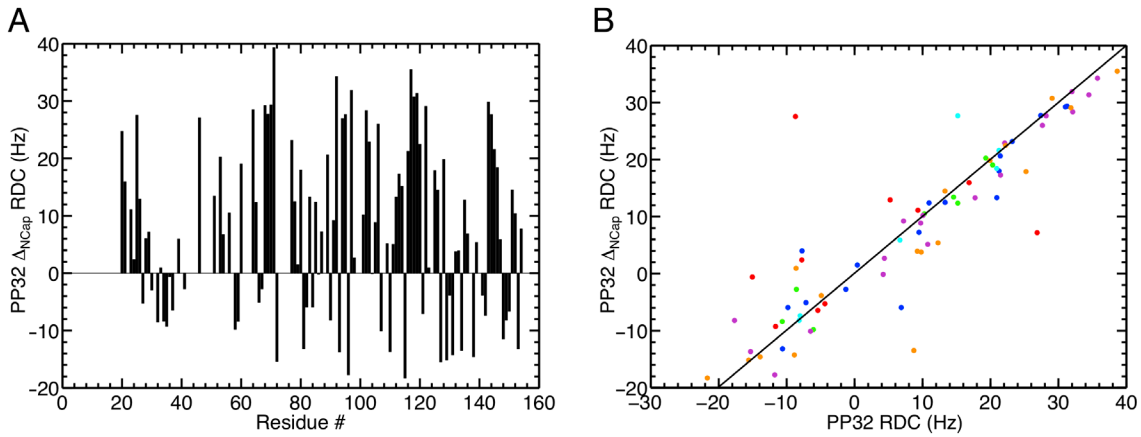


Figure 2.5. RDC of Δ_{Ncap} PP32. (A) RDC values along the sequence of Δ_{Ncap} PP32 in liquid crystalline media of 5% C12E5/Hexanol. (B) Correlation between RDCs of Δ_{Ncap} PP32 and of corresponding PP32 residues. The data are colored by repeat identity: red, 1; green, 2; blue, 3; magenta, 4; orange, 5; and cyan, C-terminal capping motif. The black line represents perfect correlation of RDCs between full-length and Δ_{Ncap} PP32.

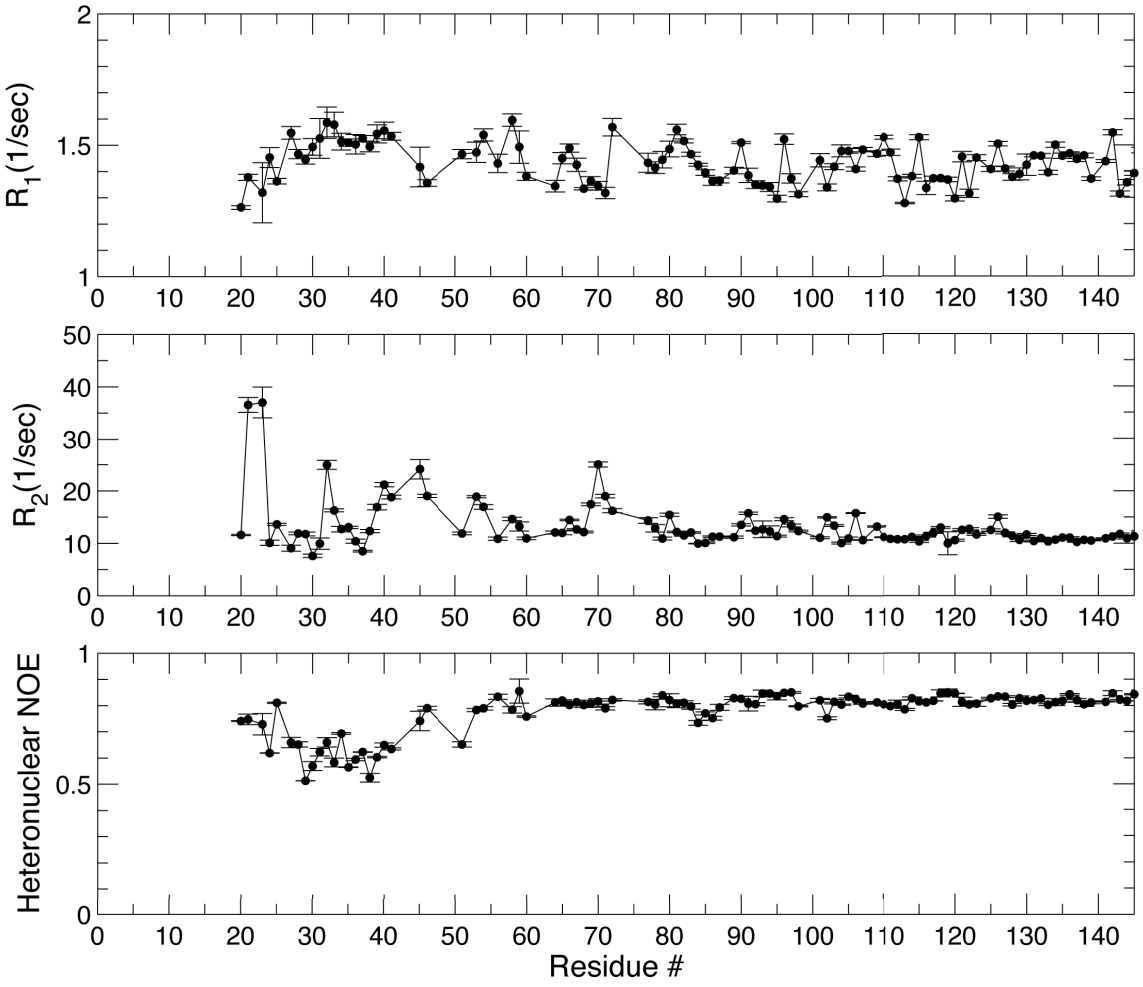


Figure 2.6. The ^{15}N spin relaxation parameters of Δ_{NCap} PP32. Residue specific R_1 (top), R_2 (middle) and heteronuclear NOE (bottom) profiles are shown with error bars. The relaxation experiments were carried out at pH 6.8, 30^{o}C.

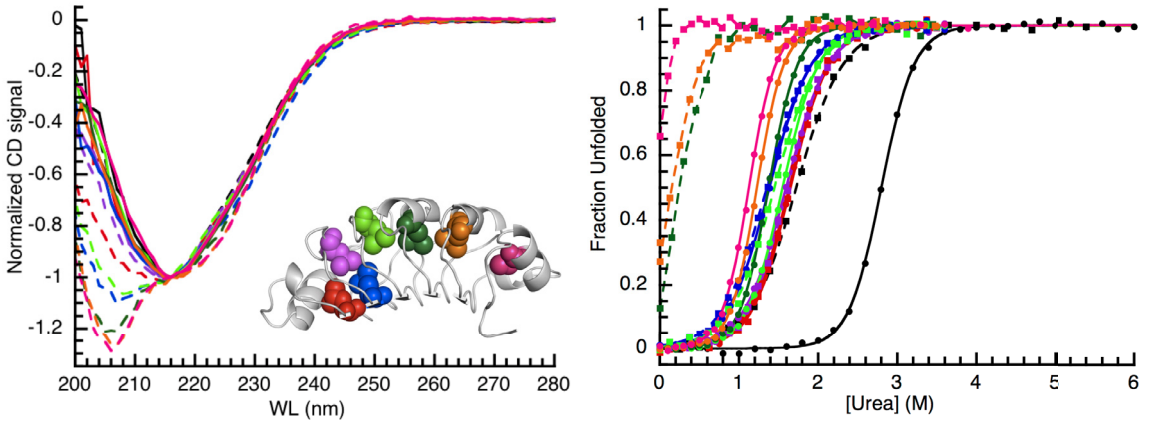


Figure 2.7. CD spectroscopy and equilibrium unfolding of core packing substitutions in PP32 and Δ_{NCap} PP32. (A) Far-UV CD shows disruption of β-sheet structure upon C-terminal packing substitutions in Δ_{NCap}, but not in full-length PP32. For direct comparison of the shape of far-UV CD spectra, the traces are normalized to -1 at 216nm. Wild-type constructs are in black. Packing variants are as depicted in the ribbon diagram: red, L22A; magenta, L37A; blue, L47A; light green, L60A; dark green, L83A; orange, L109A; pink, V135G. Full-length constructs are shown with solid lines. Δ_{NCap} constructs are shown with dashed lines. (B) Normalized urea-induced unfolding transitions monitored by far-UV CD for PP32 and variants (circles, solid line) and Δ_{NCap} PP32 and variants (squares, dashed line). Lines result from fitting a two-state unfolding model to the data. Lines for Δ_{NCap} PP32 L83A, L109A and V135G are only for guide and do not result from fitting.

Chapter III

The highly polarized C-terminal transition state of the leucine-rich repeat domain of PP32 is governed by local stability

3.1 Abstract

The leucine-rich repeat domain of PP32 is composed of five β-strand-containing repeats anchored by terminal caps. These repeats differ in sequence but are similar in structure, providing a means to connect topology, sequence and folding pathway selection. Through kinetic studies of PP32, we find folding to be rate-limited by the formation of an on-pathway intermediate. Destabilizing core substitutions reveal a transition state ensemble that is highly polarized towards the C-terminal repeat and cap. To determine if this nucleus for folding corresponds to the most stable region of PP32, we monitored amide hydrogen exchange by NMR spectroscopy. Indeed, we find the highest protection to be biased towards the C-terminus. Sequence manipulations that destabilize the C-terminus spread out the transition state towards the middle of the protein. Consistent with results for helical ankyrin repeat proteins, these results suggest that local stabilities determine folding pathways.

3.2 Introduction

Almost 50 years ago, Cyrus Levinthal suggested that protein folding is a guided process, rather than a random search (Levinthal, 1968). Since then, it has been proposed that rapid folding can be accomplished via a single preferred pathway (Kim and Baldwin, 1990; Matthews, 1993; Myers and Oas, 2002). However contrasting models of funneled, nonspecific folding capture key aspects of the folding process, including specific intermediates and transient stable structures (Oliveberg and Wolynes, 2005; Wolynes et al., 1995). To what extent does protein folding follow a single pathway, and to the extent that a single pathway dominates, what determines that pathway?

For some globular proteins, specific kinetic intermediates have been identified with features of thermodynamically stable substructures of the native state. This observation suggests specific pathways that correspond to low energy folding routes (Hoang et al., 2002; Hu et al., 2013; Jennings and Wright, 1993; McCallister et al., 2000; Nauli et al., 2001; Raschke and Marqusee, 1997). However, directly testing this thermodynamic control of folding pathway selections requires mapping of local stability. Hydrogen exchange methods have provided the most detailed energy maps of native proteins (Krishna et al., 2004), but sequence-distant contacts and irregular tertiary structures of globular proteins tend to blur the boundaries of local stability.

In contrast to globular proteins, elongated repeat proteins comprise tandem repeated secondary structural units, giving rise to regular topology lacking

sequence-distant contacts. As a result, the contributions made by different regions of a repeat protein to its folding thermodynamics and kinetics can be easily dissected and compared. If local stability dictates folding, a preferred pathway should be observed, since repeats differ in sequence. If topology drives folding, multiple parallel pathways would be observed, since the repeats are similar in structure. For designed consensus α-helical ankyrin repeat constructs, with repeats of nearly identical sequence, folding proceeds via parallel pathways (Aksel and Barrick, in press). In contrast, naturally-occurring ankyrin repeat proteins with repeats of high sequence variation fold through preferred pathways (Bradley and Barrick, 2006; Löw et al., 2007; Lowe and Itzhaki, 2007; Tang et al., 2003). Regions that initiate folding correspond to low-energy structures, based on energy landscapes determined at single repeat resolution, consistent with low-energy folding routes for α-helical repeat proteins (Bradley and Barrick, 2006; Mello and Barrick, 2004; Tripp and Barrick, 2007, 2008).

Much less is known about the folding mechanisms of β-strand-containing repeat proteins, and particularly, leucine-rich repeat (LRR) proteins. To date, the folding pathways for only two LRR proteins have been elucidated. Like the ankyrin repeat domains, the LRR domain of Internalin B (InlB) also folds through a somewhat polarized pathway, with the transition state ensemble involving the N-terminal capping motif and first three repeats (Courtemanche and Barrick, 2008a). The folding of the mRNA exporter protein TAP (TAPLRR) is more complicated, with a triangular mechanism and a diffuse transition state involving

most of the protein (Kelly et al., 2014). However, how these pathways are selected remains unknown.

To test if LRR proteins fold via low-energy pathways, we investigated the folding mechanism and stability distribution of the LRR domain of PP32. This domain is composed of a linear array of five repeats, each containing a β-strand followed by either an extended structure, a 3_{10}-helix, or a short α-helix (Figure 3.1, Huyton and Wolberger, 2007). The N- and C-termini of PP32 are capped by a helix-loop-helix motif and a β-hairpin, respectively. Deletion and mutational analysis showed that whereas the N- terminal cap contributes stability, the C-terminal cap is critical for stability: without the C-terminal cap, PP32 is entirely unfolded (Dao et al., 2014). Here we investigate the effects of destabilizing substitutions on folding kinetics and find that PP32 folds via an on-pathway intermediate with a highly polarized C-terminal pathway. We also map residue-specific stabilities throughout the LRR domain of PP32 using amide hydrogen exchange by NMR spectroscopy and find that the C-terminal repeat and cap have the highest protection factors. Thus, the region to first become structured during refolding corresponds to the most stable part of the protein. Destabilizing the C-terminus of PP32 causes the highly polarized folding pathway to become much more diffuse. These results highlight the importance of local stabilities in determining the folding mechanisms of LRR proteins, as observed with α-helical repeat proteins.

3.3 Results

Spectroscopic changes upon PP32 unfolding

To study the folding of PP32, we monitored structural changes by circular dichroism (CD) at 220 nm and by fluorescence, which monitor β-sheet structure and the tryptophan residue at the C-terminus, respectively (Figure 3.2). Urea-induced equilibrium unfolding monitored by both techniques resulted in the same C_m and $\Delta G^o_{H_2O}$ values, suggesting that CD and fluorescence capture the folding reaction. For the remainder of this work, we present data for equilibrium unfolding of PP32 and variants by CD (Figures 3.4 and 3.9B) and folding kinetics by fluorescence (Figures 3.3, 3.5 and 3.9C).

Urea dependence of folding and unfolding kinetics

To gain insights into the folding pathway of PP32, we monitored fluorescence changes upon rapid dilution of native and denatured protein to various urea concentrations by stopped flow. Both refolding and unfolding reactions appear to be multiphasic. The data are poorly fitted by a single-exponential model (gray lines, Figures 3.3A and 3.3B), resulting in large, non-random residuals (upper panels). In contrast, a double-exponential model describes the data well (black lines, Figures 3.3A and 3.3B), resulting in small, random residuals (upper panels). For refolding, a fast major phase is followed by a slow minor phase. Amplitudes for both refolding phases have the same sign, in the refolding direction. For unfolding, an early lag is followed by a slow major phase (see inset, Figure 3.3B). Amplitudes for the unfolding phases have opposite signs: the

amplitude of the major phase is in the unfolding direction whereas that for the minor phase is in the refolding direction.

The major phases for folding and unfolding both show high denaturant dependences, and join up to form a V-shaped chevron plot (circles, Figure 3.3C). Thus, these two phases are likely to report on the kinetics of the main conformational transition in PP32 folding. However, the biphasic unfolding and refolding suggest that PP32 folds via a more complicated pathway than a simple two-state kinetic mechanism.

PP32 has five prolines, each in *trans* configuration. *Cis-trans* proline isomerization reactions are known to produce a slow, denaturant-insensitive phase during refolding, with a rate constant between 0.01 and 0.1 sec^{-1} (Bradley and Barrick, 2005; Brandts et al., 1975; Courtemanche and Barrick, 2008b; Kamen and Woody, 2002a, 2002b; Kiefhaber and Schmid, 1992; Kiefhaber et al., 1990; Kloss and Barrick, 2008; Schmid et al., 1993; Schmid and Baldwin, 1978, 1979; Tang et al., 1999; Zeeb et al., 2002). The magnitude and denaturant dependence of the rate constant for the slow refolding phase in PP32 (triangles, Figure 3.3C) are consistent with proline isomerization.

The major phase of the unfolding arm shows a roll-over at 4.4 M urea. This type of nonlinearity is commonly observed for proteins folding through an intermediate (Baldwin, 1996). For PP32, an intermediate in unfolding is supported by the observation of a second kinetic phase (squares, Figure 3.3C).

The roll-over and minor phase in unfolding are consistent with a populated on-pathway kinetic intermediate (Mello et al., 2005).

To obtain stepwise rate constants and their urea dependences, we globally fit a sequential three-state model (D⇌I⇌N, see Experimental Procedures) to the major refolding rate constants, the major and minor unfolding rate constants, as well as the associated unfolding amplitudes (black lines, Figures 3.3C and 3.3D). These data are well-fitted by the model, as illustrated by the small, random residuals (Figures 3.3C and 3.3D, upper panels). Moreover, both the kinetically deduced $\Delta G^{o}_{H_2O}$ (calculated as $-RT*\ln\{k_{DI}k_{IN}/k_{NI}k_{ID}\}$) and *m*-value (calculated as $m_{DI}-m_{ID}+m_{IN}-m_{NI}$) agree well with corresponding equilibrium values (Table 3.1), strongly supporting the on-pathway intermediate mechanism for PP32 folding. This good agreement between kinetic and equilibrium values also suggests that the kinetic parameters are well determined from the fit. Although there is considerable uncertainty in the extrapolated value of the rate constant for the second step in folding (k_{IN}, Table 3.1), it is much greater than k_{DI}. Thus, formation of the intermediate from the denatured state is the rate-limiting step in folding.

Structure formation along the PP32 folding pathway

To determine the structure of PP32 along the folding pathway, we investigated the equilibrium and kinetic effects of point substitutions along the molecule, with the idea that substitutions slowing down folding kinetics are in regions that are structured in the transition states. We made substitutions in the

N- and C-terminal caps, as well as the conserved leucines on the β-strand and convex side of the repeats (Figure 3.1). Except for K137G, D146L and Y131F, all substitutions are structurally conservative, removing only part of a hydrophobic side chain, and should not introduce new interactions. All variants are destabilized by at least 1 kcal • mol^{-1} (except for K137G), but remain largely folded (Figure 3.4, Table 3.1), allowing for accurate determination of Φ-values (Sánchez and Kiefhaber, 2003).

As with WT PP32, the refolding kinetics of all variants are biphasic (Figure 3.5). The slow minor phase of each variant is similar to that of WT, consistent with it being a proline isomerization phase, rather than a folding reaction. The fast refolding phase is unaffected by substitutions from the N-terminal helical cap to repeat three (L93A), but is significantly slowed down by substitutions in the C-terminus (repeats four and five, and capping motif).

In contrast to WT PP32, for most variants (except I7A, V19A, L60A and L69A), only a single unfolding phase is observed (Figure 3.5), suggesting that the on-pathway intermediate is no longer populated to high levels. However, roll-overs at high urea concentrations persist in the chevron plots of all variants, suggesting that the intermediate continues to influence the kinetics of unfolding. Whereas the unfolding reactions of the N-terminal variants (I7A, L11A, V19A, L22A, L37A, L47A, L60A, L69A) significantly speed up, those of variants in the central repeats (L83A, L93A, L109A, L118A) are slightly faster, and those of the

C-terminal variants (Y131A, V135G, K137G, L139A, L142A, L145A, D146L) remain largely unaffected.

To quantitatively map out the folding pathway, we fit a sequential three-state model to the chevron plot of each variant (solid lines, Figure 3.5, Table 3.1). For variants that show two unfolding phases (I7A, V19A and L69A), the rate constants and amplitudes are well determined by the model. For the variants without an observed minor unfolding phase, there is not enough information to determine the full parameter set. To constrain these fits, we set m_{NI} and m_{ID} to WT values. The fitted curves describe the data accurately and the resulting parameters are well determined. Using equilibrium $\Delta G^{o}_{H_2O}$ values and the fitted k_{DI}, k_{IN}, k_{NI}, k_{ID} values for WT PP32 and variants in the absence of urea, we calculated Φ-values for TS1 (D→I), the on-pathway intermediate, and TS2 (I→N). The Φ-value reflects the extent to which a substitution affects a rate constant relative to a related equilibrium constant (Matthews, 1987; Otzen et al., 1994). A low Φ-value indicates that the site of substitution is not structured at a specific stage in folding (e.g., TS1, I, TS2), whereas a high Φ-value indicates high structure. For TS1, which is formed in the rate-limiting step, Φ-values are low in the N-terminal cap and first three repeats, slightly increase in the fourth repeat, and drastically increase in the fifth repeat and C-terminal cap (Figure 3.6A). At residue 146, the most C-terminal residue we substituted, the Φ-value decreases to an intermediate level. These Φ-values suggest that the C-terminal repeat and cap are structured in TS1, while the rest of the molecule remains largely unfolded

(Figure 3.6B). For the intermediate state, Φ-values in the N-terminal repeats increase to moderate values, indicating partial structure formation. For TS2, Φ-values are high at most positions, indicating structure formation along the entire PP32. Only the residues at the very beginning of the N-terminal cap appear to become structured after TS2.

If PP32 folds via a polarized C-terminal pathway, removing the N-terminal region should not affect the first step of folding. To test this prediction, we measured folding kinetics of a PP32 construct that lacks the N-terminal α-helical capping motif (Δ_{NCap} PP32, Dao et al., 2014). The urea-dependent refolding phases of Δ_{NCap} PP32 are similar to those of WT PP32 (Figure 3.5), whereas the slow unfolding phase significantly speeds up. These observations are consistent the results from Φ-value analysis, confirming that the N-terminal cap is not involved in the first step of folding.

Residue-specific hydrogen exchange of PP32

To map out the distribution of stability along PP32, we monitored exchange rates of backbone amide hydrogens with solvent using NMR. Hydrogen exchange (HX) provides residue-specific information on the distribution of stability within the native state (Krishna et al., 2004). The amide hydrogens in PP32 exchange over very different timescales: many peaks disappear prior to acquisition of the first spectrum, whereas others persist for over three months (Figures 3.7A and 3.7B). For peaks with measurable exchange profiles, rate constants for exchange (k_{ex}) were fitted from peak heights, and residue-specific

protection factors (PFs) were calculated as the ratio of the observed k_{ex} values to those for exchange from fully solvent-exposed unstructured state (k_{int}, see Experimental Procedures). A plot of the PFs as a function of sequence reveals increasing protection from the N-terminus to the C-terminus (Figure 3.7C). This trend (at pD 6.7) is also observed for pD 6.2 and pD 7.2 (data not shown), indicating that exchange is limited by stability, and not by opening kinetics. Therefore, the C-terminal region appears to have much a higher local stability than the N-terminal region, consistent with our previous findings that PP32 completely unfolds upon removal of part of the C-terminal cap, but retains C-terminal structure upon removal the N-terminal α-helical cap (Dao et. al., 2014).

Transition state structure can be directly mapped to local stability by comparing Φ-values with PFs at the same sites (Figure 3.8). The amide hydrogens of these residues are buried or involved in hydrogen bonds. Moreover, the two substituted leucines within each repeat are structurally conserved, providing direct comparison of different regions of PP32. The stability of PP32 increases gradually towards the C-terminus and is highest for repeats four and five and part of the C-terminal cap. Though the stability map is more spread out than the Φ-value map, both are polarized towards the C-terminus, suggesting that the folding pathway of PP32 is determined by local energetics.

Consequences of destabilizing the C-terminus on the folding pathway of PP32

To further test if stability dictates the location of the transition state ensemble, we determined the folding pathway of a C-terminally destabilized PP32 variant,

Y131F/D146L (YD). Since the structure of YD is highly similar to that of WT PP32 (Dao et. al., 2014), the destabilizing effect should be localized to the C-terminus. Though YD is significantly destabilized, it retains sufficient stability for Φ-value analysis (Figure 3.4, Table 3.1). For direct comparison, we made the same substitutions along the YD construct as we made in WT PP32 (Figure 3.9A). The YD variants are highly destabilized, to about the same extent as in the WT background. Complete unfolding transitions are observed for all but two variants (L93A and L139A; Figure 3.9B).

To determine the effects of the substitutions on folding kinetics as the C-terminus is destabilized, we compare the major phases of the chevron plots for the variants in the WT and YD backgrounds (Figure 3.9C). In the YD background, we observe increased sensitivity to substitutions at the N-terminus and decreased substitutions at the C-terminus, compared to WT PP32. To quantitatively assess the effects of destabilizing the C-terminus on the structure of transition state ensemble for the rate-limiting step in refolding of PP32, we compare Φ-values for the five variants in the YD and WT backgrounds (Figure 3.9D). The Φ-values for L139A are the highest in both backgrounds, indicating that both fold via their C-termini. However, whereas Φ-values in WT PP32 show a highly polarized transition state restricted to repeat five, the significant positive Φ-values in repeat three and four of YD indicate a more dispersed transition state. Thus, destabilizing the C-terminus (the preferred folding route in WT PP32)

spreads out the transition state, switching from a highly polarized structure to a more extended region.

3.4 Discussion

The linear LRR domain of PP32 is well suited for mechanistic studies of folding. It has sufficient stability to accommodate destabilizing point substitutions, permitting Φ-values to be determined reliably (Figure 3.2, Table 3.1). Its linear, repetitive architecture allows the results of Φ-values analysis to be evaluated in the context of local stability variation in structurally conserved fragments.

Overall folding mechanism of PP32

Kinetic analysis of PP32 refolding and unfolding reveals an on-pathway intermediate. On-pathway intermediates are also observed for the α-helical ankyrin repeat domains including Notch (Bradley and Barrick, 2006), p16 (Tang et al., 1999), and IκBα (DeVries et al., 2011), but not for the two LRR proteins InlB (Courtemanche and Barrick, 2008b) and TAPLRR (Kelly et al., 2014) studied to date. For PP32, formation of this intermediate from the denatured state is rate limiting in refolding at low to moderate urea concentrations. Although the rate constant for the I to N step (k_{IN}) cannot be determined with great precision, given the large extrapolation in the chevron plot, it is clear that k_{IN} exceeds the rate constant for the D to I step (k_{DI}) by several orders of magnitude (Table 3.1), and may be as large as estimated "speed limits" for folding (Kubelka et al., 2004; Yang and Gruebele, 2003).

In contrast, the rate constants for the two unfolding steps (k_{NI} and k_{ID}) are similar to each other at all urea concentrations, with the first unfolding step (from N to I) about five-fold faster than the second (from I to D) (Table 3.1, Figure

3.3C). The similarity of these two rate constants produces a lag in unfolding at high denaturant concentrations, where I is more stable than N, and thus becomes transiently populated during unfolding. Below 4.4 molar urea, I is less stable than N; thus, unfolding shifts to a single exponential conversion from N to D, but is rate-limited by the transition state between I and D. As a result, at the urea midpoint for the reaction between N and I (see minimum in minor phase chevron, Figure 3.3C), there is a roll-over in the unfolding arm of the chevron for the major phase. This roll-over reflects a shift from a modest denaturant dependence above this midpoint (resulting from surface area changes between I and TS1, Figure 3.6B) to a more severe denaturant dependence below this midpoint (resulting from surface area changes between N and TS1).

The discrete C-terminal folding pathway is selected based on local stability

Φ-value analysis for 20 substitutions along PP32 reveals a highly polarized rate-limiting transition state (D to I) involving only the C-terminal repeat and cap (Figures 3.4-5, Table 3.1). Kinetic studies of a construct without the N-terminal cap (Δ_{NCap} PP32) further support the C-terminal polarized folding pathway of PP32. The degree to which the folding nucleus of PP32 is localized is striking in comparison to other naturally occurring repeat proteins, such as InlB and the ankyrin domains of Notch, p16 and Iκbα. Structure in the rate-limiting transitions states for these proteins includes roughly half of each protein (Bradley and Barrick, 2006; Courtemanche and Barrick, 2008a; DeVries et al., 2011; Tang et al., 2003).

Through amide hydrogen exchange measurements monitored by NMR, we have found that PP32 is more stable towards the C-terminus (Figure 3.7). The correspondence between local stability and high Φ-values is consistent with stability dictating folding pathway selection. However, whereas the Φ-values for the TS1 are about zero for most of the protein and almost one for only residues 135-142, indicative of a highly polarized pathway, the stability distribution determined by HX is more spread out (Figure 3.8). This difference results from the fact that native-state hydrogen exchange and Φ-value analysis probe different parts of the energy landscape. Φ-value analysis probes the minimum structural unit required to lead to downhill folding from the denatured state. In contrast, HX studies observe unfolding reactions in the context of the folded state, such that structural modules (foldons, Englander et al., 2007) can stabilize one another. Work on consensus α-helical repeat proteins (Aksel et al., 2011; Wetzel et al., 2008) and bacterial LRR proteins (TPD and DB, unpublished data) have shown that the interfacial interactions between repeats are strongly stabilizing. Thus, we expect the highly stable C-terminus of PP32 to protect neighboring repeats from exchange, broadening the observed stability map toward the N-terminus.

Previous studies have shown that transition state structures can be shifted by sequence substitutions, consistent with pathway selection by local stability. For the four-ankyrin repeat protein myotrophin, destabilizing C-terminal substitutions shift the folding pathway from the C-terminus to the N-terminus

(Lowe and Itzhaki, 2007). In contrast, a set of consensus substitutions stabilizing the last two repeats of the Notch ankyrin domain shifts the folding pathway from the middle repeats to the C-terminal repeats (Tripp and Barrick, 2007, 2008). Here, we show that the C-terminally destabilized variant Y131F/D146L folds via a much more diffuse transition state than WT PP32 (Figure 3.9), providing additional evidence that folding pathways are determined by local energetics. However, unlike the ankyrin domains, C-terminal repeats of PP32 still participate in the transition state despite local destabilization, consistent with the C-terminus being significantly more stable than the rest of PP32. To reroute the folding pathway of PP32 so that the C-terminal repeats fold after the transition state ensemble, a combination of substitutions destabilizing the C-terminus and stabilizing the N-terminal repeats may be required.

Possible determinants for folding pathway selections in other LRR proteins

The folding mechanisms of three LRR proteins whose pathways have been elucidated to date are very different from one another. Like PP32, the structure of the transition state ensemble for InlB is also localized, although to the N-terminus instead of the C-terminus (Courtemanche and Barrick, 2008a). In contrast, TAPLRR folds via a diffuse transition state in which the first three (of four) repeats are structured and the N-terminal cap makes non-native contacts with the first repeat (Kelly et al., 2014). It is possible that pathway selections for both InlB and TAPLRR are also based on local energetics, as observed for PP32, in

which case the N-terminal capping motif and first three repeats would be the most stable region of InlB, whereas the repeats of TAPLRR would have more uniform stabilities.

Secondary structure might also contribute to the variations in folding pathways among these three LRR proteins. α-helices have been suggested to fold faster than β-stranded structure (Plaxco et al., 2000). Although PP32, InlB and TAPLRR all have β-sheets spanning their concave surfaces, they differ significantly in helix content. The C-terminal LRRs of PP32 have α-helices on their convex surfaces (Figure 3.1A), but the N-terminal LRRs do not. Thus, these helices may help initiate folding of the C-terminal LRRs. The transition state for folding of InlB includes the N-terminal cap, which is the only region with α-helices. The uniform distribution of α-helices along the TAPLRR domain may contribute to its diffuse transition state and may lead to parallel folding pathways, as recently reported for an α-helical ankyrin repeat protein series **(**Aksel and Barrick, in press**).** However, helical structure alone is not enough to determine folding routes since preferred pathways are observed for α-helical repeats proteins, where repeats all have a high helix content (Bradley and Barrick, 2006; Löw et al., 2007; Lowe and Itzhaki, 2007; Tang et al., 2003; Tripp and Barrick, 2008).

Biological implications

Both the asymmetric stability distribution and highly polarized C-terminal folding pathway have implications for folding and fluctuation in cellular syntheses

and transport. We have previously shown that the C-terminally truncated PP32 is unfolded in solution (Dao, et. al., 2014). Hence, in addition to providing a kinetically favored route to folding, the C-terminus is required to maintain the native fold. Thus, PP32 is not expected to fold cotranslationally. If cotranslational folding were important to avoid aggregation/misfolding (Clark, 2004), a more N-terminal folding pathway (and stability distribution) would be expected. It is plausible that the ribosome could significantly reprogram the folding landscape, transiently stabilizing the N-terminal repeats. Although such a shift seems unlikely, given the strong polarization found here, identification of such a shift would be a compelling demonstration of the importance of cotranslational folding.

PP32 is primarily a nuclear protein. Nuclear import is mediated by a nuclear localization signal in its acidic region, C-terminal to the LRR domain (Matsubae et al., 2000). However, PP32 can also shuttle back to the cytoplasm through binding of the LRR domain to nuclear export protein Crm1 (Brennan et al., 2000). Crm1 recognizes its cargo through leucine-rich nuclear export signals (NES) with a consensus sequence of $x\text{-}\Phi^{0}\text{-}x_{2}\text{-}\Phi^{1}\text{-}x_{2,3}\text{-}\Phi^{2}\text{-}x_{2,3}\text{-}\Phi^{3}\text{-}x\text{-}\Phi^{4}$, where hydrophobic residues Φ (most commonly leucine) are separated by one to three residues x (often polar or charged; Güttler et al., 2010; Xu et al., 2012). This consensus sequence is found in PP32 LRRs 1-4, at least one of which is likely to interact with Crm1. Structures of Crm1 complexes have shown that the conserved Φ's of the consensus sequence dock into the pockets of the hydrophobic cleft of Crm1 (Dong et al., 2009; Monecke et al., 2013). As the conserved Φ's of PP32 make

up its core and are not solvent exposed, binding to Crm1 is likely to require partial unfolding of PP32. Our HX result showing that the N-terminus is much less stable than the C-terminus suggests the N-terminal repeats (LRRs 1-3) as likely binding motifs for Crm1.

3.5 Materials and Methods

Subcloning, protein expression, and purification

The gene encoding PP32 was a kind gift from the laboratory of Dr. Cynthia Wolberger. Constructs encoding PP32, Δ_{NCap} PP32 and PP32 variants I7A, V19A, L69A, L93A, L139A, Y131F, D146L and Y131F/D146L have been describe previously (Dao et al., 2014). Additional point substitutions were made using Quikchange (Stratagene, La Jolla, CA). Protein were expressed and purified as described (Dao et al., 2014).

Fluorescence spectroscopy

Fluorescence emission spectra were collected on an Aviv ATF 105 spectropolarimeter (Lakewood, NJ), in a 1.0 cm path-length cuvette. The protein concentration was 3 μM, in storage buffer (20 mM $NaPO_4$, 150 mM NaCl, 0.1 mM TCEP, pH 7.8) with 0 or 5.4 M urea, at 20 °C.

Equilibrium unfolding

Urea-induced unfolding was monitored by CD (Aviv Model 400 CD spectrometer, Lakewood, NJ) at 220 nm or fluorescence by exciting at 295 nm and recording emission at 324 nm. Urea was deionized by chromatography over mixed-bed resin (Bio-Rad, Hercules, CA). Urea concentration was determined by refractometry. Urea titrations were carried out using a computer-controlled Microlab syringe titrator (Hamilton, Reno, NV). Samples contained 2-4 μM protein, 20 mM sodium phosphate, 150 mM NaCl, 0.1 mM TCEP (pH 7.8). At

each urea concentration, samples were equilibrated for 5 minutes at 20 ^{o}C and signal averaged for 30 seconds. Two-state analysis of equilibrium unfolding transitions were carried out as described by Street et al. (2008).

Kinetic folding studies

Fluorescence-detected unfolding and refolding kinetic measurements were made on an Applied Photophysics SX. 18MV-R stopped-flow rapid mixing device (Leatherhead, England). Emission was monitored using a 320 cutoff filter, following excitation at 280 nm to monitor changes in the environment surrounding the single tryptophan at the C-terminus. Final protein concentrations were 1 to 3 μM. Experiments were done in 20 mM $NaPO_4$, 150 mM NaCl, 0.1 mM TCEP (pH 7.8) and at 20 ^{o}C. Unfolding and refolding amplitudes and rate constants were determined using non-linear least-squares to fit the following equation to the individual progress curves:

$$Y_{obs} = Y_{\infty} + \sum_{i} \Delta Y_i \exp^{-k_i t}$$

Y_{∞} represents the fluorescence at equilibrium. ΔY_i and k_i represent the change in fluorescence signal contributed by the i^{th} phase and the rate constant for the i^{th} phase, respectively. Two phases were necessary and sufficient to describe PP32 refolding and unfolding kinetics of WT PP32 and variants I7A, V19A, L60A and L69A. Only one phase was sufficient to describe the unfolding for the rest of the variants.

To capture this complexity in refolding and unfolding, data were fitted with a sequential three-state model:

$$D \underset{k_{DI}}{\overset{k_{DI}}{\rightleftharpoons}} I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftharpoons}} N$$

where the denatured (D) state is converted to the native (N) state through a single on-pathway intermediate (I) (Khorasanizadeh et al., 1996; Sánchez and Kiefhaber, 2003, 2003). Microscopic rate constants from the sequential three-state model and their urea dependences were obtained from the global fitting of the observed rate constants for refolding and unfolding and, when applicable, the associated amplitudes for major and minor unfolding phases, all as a function of urea. The program Profit 6.1.16 (Quantum Soft, Switzerland) was used for fitting, with scripts kindly provided by by Drs. Thomas Kiefhaber and Andreas Moeglich.

Φ-values were calculated from the following relationships:

$$\Phi_{TS1} = \frac{RTln(k^{mut}_{DI,H_2O}/k^{wt}_{DI,H_2O})}{\Delta G^{o\,mut}_{H_2O} - \Delta G^{o\,wt}_{H_2O}}$$

$$\Phi_{I} = \frac{RT\ln[(k^{mut}_{DI,H_2O}/k^{mut}_{ID,H_2O})/(k^{wt}_{DI,H_2O}/k^{wt}_{ID,H_2O})}{\Delta G^{o\,mut}_{H_2O} - \Delta G^{o\,wt}_{H_2O}}$$

$$\Phi_{TS2} = \frac{RT\ln[(k^{mut}_{DI,H_2O}/k^{mut}_{ID,H_2O} * k^{mut}_{IN,H_2O})/(k^{wt}_{DI,H_2O}/k^{wt}_{ID,H_2O} * k^{wt}_{IN,H_2O})}{\Delta G^{o\,mut}_{H_2O} - \Delta G^{o\,wt}_{H_2O}}$$

where the rate constants and free energy changes were extrapolated to zero molar denaturant and *mut* and *wt* indicate values for variants and WT PP32, respectively. Free energies of unfolding from N to D were determined from equilibrium urea unfolding experiments.

Hydrogen exchange of WT PP32

^{15}N-labeled PP32 (800 μM in 20 mM sodium phosphate, 50 mM NaCl, 0.2 mM TCEP, pH 6.8) was exchanged into D_2O buffer (20 mM sodium phosphate, 50 mM NaCl, 0.1 mM EDTA, 0.2 mM TCEP, pDs 6.2, 6.7 or 7.2) by centrifugal gel filtration as described by Jeng and Englander (1991). 2 mL of preswollen Sephadex G-25 Fine (GE Healthcare) was placed into a 3-mL spin column (Fisher Scientific), washed 4-5 times with 2 mL of D_2O buffer by low-speed spinning (table top swing bucket centrifuge, for 3 min at 3000 x g), followed by 560 mL of protein sample. Final protein concentration was about 650 μM in 95% D_2O sample buffer.

^{1}H-^{15}N HSQC spectra were immediately collected after solvent exchange on a Bruker 600 MHz spectrometer equipped with a cryoprobe at 20 ^{o}C. About 20 spectra were collected during the first 12 hours. Subsequently, one spectrum was collected every few hours, everyday, then every few days for the next three months. Spectra were processed using NMRPipe (Delaglio et al., 1995), and displayed and analyzed with Sparky (Goddard, T. D. and Kneller, D. G.). Assignments for amide resonances of PP32 are from Dao et al., 2014. Peak heights were determined in Sparky. To compare the exchange rates of different residues of PP32, we calculated protection factors (*PF*) with the following equation:

$$PF = \frac{k_{\text{int}}}{k_{\text{ex}}}$$

where k_{int} is the rate constant for exchange from amino acids in random coil (Bai

et al., 1993) and k_{ex} is the observed exchange rate constant obtained from fitting a single- exponential decay function to the change in the heights of amide cross-peaks in the series of HSQC spectra.

Acknowledgements

We thank Drs. Cythia Wolberger for providing us with the gene encoding the LRR domain PP32, and Carlos A. Castaneda for assistance in analysis of HX data. We thank the JHU Biomolecular NMR Center for providing the facilities and resources. This research is supported by NIH grant GM068462 to D.B.

3.6 References

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. (1993). Primary Structure Effects on Peptide Group Hydrogen Exchange. Proteins *17*, 75–86.

Baldwin, R.L. (1996). On-pathway versus off-pathway folding intermediates. Fold. Des. *1*, R1–R8.

Bradley, C.M., and Barrick, D. (2005). Effect of multiple prolyl isomerization reactions on the stability and folding kinetics of the notch ankyrin domain: experiment and theory. J. Mol. Biol. *352*, 253–265.

Bradley, C.M., and Barrick, D. (2006). The Notch Ankyrin Domain Folds via a Discrete, Centralized Pathway. Structure *14*, 1303–1312.

Brandts, J.F., Halvorson, H.R., and Brennan, M. (1975). Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. Biochemistry (Mosc.) *14*, 4953–4963.

Brennan, C.M., Gallouzi, I.-E., and Steitz, J.A. (2000). Protein Ligands to Hur Modulate Its Interaction with Target Mrnas in Vivo. J. Cell Biol. *151*, 1–14.

Clark, P.L. (2004). Protein folding in the cell: reshaping the folding funnel. Trends Biochem. Sci. *29*, 527–534.

Courtemanche, N., and Barrick, D. (2008a). The leucine-rich repeat domain of Internalin B folds along a polarized N-terminal pathway. Struct. Lond. Engl. 1993 *16*, 705–714.

Courtemanche, N., and Barrick, D. (2008b). Folding thermodynamics and kinetics of the leucine-rich repeat domain of the virulence factor Internalin B. Protein Sci. Publ. Protein Soc. *17*, 43–53.

Dao, T.P., Majumdar, A., and Barrick, D. (2014). Capping motifs stabilize the LRR protein PP32 and rigidify adjacent repeats. Protein Sci. Publ. Protein Soc.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

DeVries, I., Ferreiro, D.U., Sánchez, I.E., and Komives, E.A. (2011). Folding kinetics of the cooperatively folded subdomain of the IκBα ankyrin repeat domain. J. Mol. Biol. *408*, 163–176.

Dong, X., Biswas, A., Süel, K.E., Jackson, L.K., Martinez, R., Gu, H., and Chook, Y.M. (2009). Structural basis for leucine-rich nuclear export signal recognition by CRM1. Nature *458*, 1136–1141.

Englander, S.W., Mayne, L., and Krishna, M.M.G. (2007). Protein folding and misfolding: mechanism and principles. Q. Rev. Biophys. *40*, 287–326.

Goddard, T. D., and Kneller, D. G. SPARKY J.

Güttler, T., Madl, T., Neumann, P., Deichsel, D., Corsini, L., Monecke, T., Ficner, R., Sattler, M., and Görlich, D. (2010). NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. Nat. Struct. Mol. Biol. *17*, 1367–1376.

Hoang, L., Bédard, S., Krishna, M.M.G., Lin, Y., and Englander, S.W. (2002). Cytochrome c folding pathway: Kinetic native-state hydrogen exchange. Proc. Natl. Acad. Sci. *99*, 12173–12178.

Hu, W., Walters, B.T., Kan, Z.-Y., Mayne, L., Rosen, L.E., Marqusee, S., and Englander, S.W. (2013). Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. Proc. Natl. Acad. Sci. *110*, 7684–7689.

Jeng, M.-F., and Englander, S.W. (1991). Stable submolecular folding units in a non-compact form of cytochrome c. J. Mol. Biol. *221*, 1045–1061.

Jennings, P.A., and Wright, P.E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science *262*, 892–896.

Kamen, D.E., and Woody, R.W. (2002a). Folding Kinetics of the Protein Pectate Lyase C Reveal Fast-Forming Intermediates and Slow Proline Isomerization†,‡. Biochemistry (Mosc.) *41*, 4713–4723.

Kamen, D.E., and Woody, R.W. (2002b). Identification of proline residues responsible for the slow folding kinetics in pectate lyase C by mutagenesis. Biochemistry (Mosc.) *41*, 4724–4732.

Kelly, S.E., Meisl, G., Rowling, P.J.E., McLaughlin, S.H., Knowles, T., and Itzhaki, L.S. (2014). Diffuse transition state structure for the unfolding of a leucine-rich repeat protein. Phys. Chem. Chem. Phys.

Khorasanizadeh, S., Peters, I.D., and Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nat. Struct. Mol. Biol. *3*, 193–205.

Kiefhaber, T., and Schmid, F.X. (1992). Kinetic coupling between protein folding and prolyl isomerization: II. Folding of ribonuclease A and ribonuclease T1. J. Mol. Biol. *224*, 231–240.

Kiefhaber, T., Quaas, R., Hahn, U., and Schmid, F.X. (1990). Folding of ribonuclease T1. 1. Existence of multiple unfolded states created by proline isomerization. Biochemistry (Mosc.) *29*, 3053–3061.

Kloss, E., and Barrick, D. (2008). Thermodynamics, Kinetics, and Salt dependence of Folding of YopM, a Large Leucine-rich Repeat Protein. J. Mol. Biol. *383*, 1195–1209.

Krishna, M.M.G., Hoang, L., Lin, Y., and Englander, S.W. (2004). Hydrogen exchange methods to study protein folding. Methods *34*, 51–64.

Kubelka, J., Hofrichter, J., and Eaton, W.A. (2004). The protein folding "speed limit."Curr. Opin. Struct. Biol. *14*, 76–88.

Löw, C., Weininger, U., Zeeb, M., Zhang, W., Laue, E.D., Schmid, F.X., and Balbach, J. (2007). Folding Mechanism of an Ankyrin Repeat Protein: Scaffold and Active Site Formation of Human CDK Inhibitor p19INK4d. J. Mol. Biol. *373*, 219–231.

Lowe, A.R., and Itzhaki, L.S. (2007). Rational redesign of the folding pathway of a modular protein. Proc. Natl. Acad. Sci. *104*, 2679–2684.

Matsubae, M., Kurihara, T., Tachibana, T., Imamoto, N., and Yoneda, Y. (2000). Characterization of the nuclear transport of a novel leucine-rich acidic nuclear protein-like protein. FEBS Lett. *468*, 171–175.

McCallister, E.L., Alm, E., and Baker, D. (2000). Critical role of β-hairpin formation in protein G folding. Nat. Struct. Mol. Biol. *7*, 669–673.

Mello, C.C., and Barrick, D. (2004). An experimentally determined protein folding energy landscape. Proc. Natl. Acad. Sci. U. S. A. *101*, 14102–14107.

Mello, C.C., Bradley, C.M., Tripp, K.W., and Barrick, D. (2005). Experimental Characterization of the Folding Kinetics of the Notch Ankyrin Domain. J. Mol. Biol. *352*, 266–281.

Monecke, T., Haselbach, D., Voß, B., Russek, A., Neumann, P., Thomson, E., Hurt, E., Zachariae, U., Stark, H., Grubmüller, H., et al. (2013). Structural basis

for cooperativity of CRM1 export complex formation. Proc. Natl. Acad. Sci. *110*, 960–965.

Nauli, S., Kuhlman, B., and Baker, D. (2001). Computer-based redesign of a protein folding pathway. Nat. Struct. Mol. Biol. *8*, 602–605.

Oliveberg, M., and Wolynes, P.G. (2005). The experimental survey of protein-folding energy landscapes. Q. Rev. Biophys. *38*, 245–288.

Plaxco, K.W., Simons, K.T., Ruczinski, I., and Baker, D. (2000). Topology, Stability, Sequence, and Length: Defining the Determinants of Two-State Protein Folding Kinetics. Biochemistry (Mosc.) *39*, 11177–11183.

Raschke, T.M., and Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. Nat. Struct. Mol. Biol. *4*, 298–304.

Sánchez, I.E., and Kiefhaber, T. (2003). Evidence for Sequential Barriers and Obligatory Intermediates in Apparent Two-state Protein Folding. J. Mol. Biol. *325*, 367–376.

Schmid, F.X., and Baldwin, R.L. (1978). Acid catalysis of the formation of the slow-folding species of RNase A: Evidence that the reaction is proline isomerization. Proc. Natl. Acad. Sci. U. S. A. *75*, 4764–4768.

Schmid, F.X., and Baldwin, R.L. (1979). The rate of interconversion between the two unfolded forms of ribonuclease A does not depend on guanidinium chloride concentration. J. Mol. Biol. *133*, 285–287.

Schmid, F.X., Mayr, L.M., Mücke, M., and Schönbrunner, E.R. (1993). Prolyl isomerases: role in protein folding. Adv. Protein Chem. *44*, 25–66.

Street, T.O., Courtemanche, N., and Barrick, D. (2008). Protein Folding and Stability Using Denaturants. In Methods in Cell Biology, I. Dr. John J. Correia and Dr. H. William Detrich, ed. (Academic Press), pp. 295–325.

Tang, K.S., Guralnick, B.J., Wang, W.K., Fersht, A.R., and Itzhaki, L.S. (1999). Stability and folding of the tumour suppressor protein p16. J. Mol. Biol. *285*, 1869–1886.

Tang, K.S., Fersht, A.R., and Itzhaki, L.S. (2003). Sequential Unfolding of Ankyrin Repeats in Tumor Suppressor p16. Structure *11*, 67–73.

Tripp, K.W., and Barrick, D. (2007). Enhancing the Stability and Folding Rate of a Repeat Protein through the Addition of Consensus Repeats. J. Mol. Biol. *365*, 1187–1200.

Tripp, K.W., and Barrick, D. (2008). Rerouting the folding pathway of the Notch ankyrin domain by reshaping the energy landscape. J. Am. Chem. Soc. *130*, 5681–5688.

Wolynes, P.G., Onuchic, J.N., and Thirumalai, D. (1995). Navigating the folding routes. Science *267*, 1619–1620.

Xu, D., Farmer, A., Collett, G., Grishin, N.V., and Chook, Y.M. (2012). Sequence and structural analyses of nuclear export signals in the NESdb database. Mol. Biol. Cell *23*, 3677–3693.

Yang, W.Y., and Gruebele, M. (2003). Folding at the speed limit. Nature *423*, 193–197.

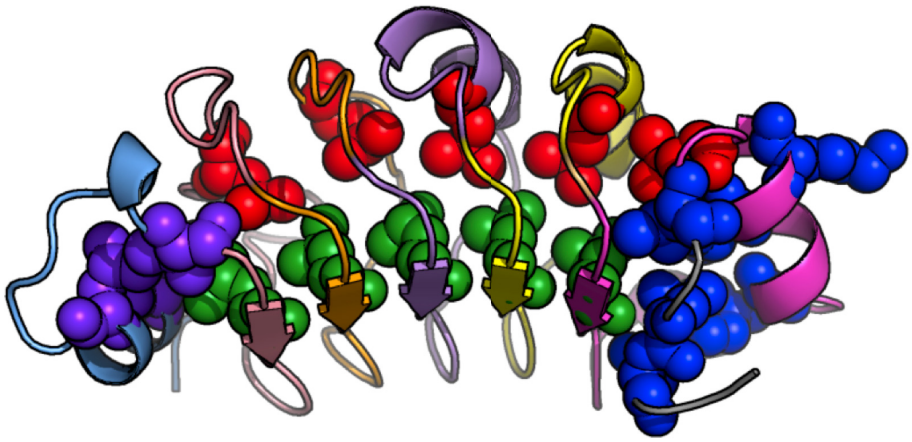
Zeeb, M., Rösner, H., Zeslawski, W., Canet, D., Holak, T.A., and Balbach, J. (2002). Protein folding and stability of human CDK inhibitor p19INK4d. J. Mol. Biol. *315*, 447–457.

Table 3.1. Fitted parameters for two-state equilibrium unfolding and three-state kinetic refolding and unfolding PP32 variants

	k_{DI,H_2O}	k_{ID,H_2O}	k_{IN,H_2O}	k_{NI,H_2O}	m_{DI}	m_{ID}	m_{IN}	m_{NI}	$\Delta G^{o}_{H_2O,eq}$	$\Delta G^{o}_{H_2O,kin}$	m_{eq}	m_{kin}
Wild-type	5.60	0.53	$3.47x10^5$	2.40	-0.88	0.36[a]	-1.36	0.35[a]	7.93±0.18	8.29±0.01	2.86±0.02	2.95±0.02
I7A	6.38	0.80	$3.08x10^5$	36	-0.88	0.45[a]	-1.48	0.16[a]	5.86±0.06	6.48±1.87	2.85±0.04	2.97±0.04
L11A	6.93	2.01	$3.83x10^4$	4.78	-0.91	0.36[b]	-1.22	0.35[b]	5.92±0.10	5.95±0.04	2.95±0.04	2.85±0.02
V19A	6.55	1.55	$2.53x10^5$	4.57	-0.89	0.37[a]	-1.52	0.33[a]	6.69±0.17	7.20±0.06	3.05±0.13	3.11±0.02
L22A	6.99	2.52	$6.66x10^2$	2.51	-0.92	0.36[b]	-0.72	0.35[b]	3.88±0.05	3.84±0.01	2.36±0.03	2.34±0.02
L37A	6.52	2.76	$5.35x10^2$	2.23	-0.93	0.36[b]	-0.72	0.35[b]	3.68±0.15	3.69±0.08	2.28±0.07	2.36±0.11
L47A	7.51	3.27	$5.52x10^2$	1.68	-0.87	0.36[b]	-1.00	0.35[b]	2.98±0.17	3.86±0.06	2.22±0.06	2.58±0.13
L60A	5.79	0.49	$6.61x10^2$	0.88	-0.80	0.36[b]	-1.79	0.35[b]	4.06±0.10	5.29±0.01	2.61±0.01	3.31±0.01
L69A	6.28	0.79	$3.20x10^3$	0.29	-0.88	0.60[a]	-1.74	0.57[a]	4.99±0.13	6.63±0.01	3.02±0.03	3.79±0.02
L83A	9.40	0.97	$8.40x10^1$	0.89	-1.39	0.36[b]	-0.83	0.35[b]	4.46±0.02	3.97±0.01	3.25±0.02	2.93±0.04
L93A	5.26	0.88	$6.07x10^2$	0.98	-1.11	0.36[b]	-1.41	0.35[b]	4.95±0.19	4.79±0.04	3.47±0.11	3.23±0.18
L109A	0.89	0.65	$5.23x10^2$	0.57	-1.13	0.36[b]	-1.35	0.35[b]	4.56±0.08	4.15±0.01	3.37±0.17	3.18±0.02
L118A	0.71	1.07	$3.13x10^2$	1.14	-1.48	0.36[b]	-0.85	0.35[b]	3.60±0.12	3.03±0.03	3.26±0.12	3.03±0.03
Y131F	12.68	0.64	$2.93x10^4$	2.82	-1.02	0.36[b]	-1.30	0.35[b]	6.97±0.09	7.13±0.03	3.05±0.04	3.03±0.06
V135G	0.03	0.65	$3.29x10^4$	1.12	-1.21	0.36[b]	-1.46	0.35[b]	4.07±0.19	4.13±0.60	3.61±0.16	3.38±0.04
K137G	1.27	0.57	$6.72x10^5$	5.41	-0.92	0.36[b]	-1.39	0.35[b]	7.26±0.12	7.30±1.24	3.04±0.17	3.02±0.03
L139A	0.03	0.66	$1.01x10^4$	0.64	-0.77	0.36[b]	-1.45	0.35[b]	4.40±0.16	3.85±0.24	3.36±0.11	2.93±0.03
L142A	0.03	0.67	$9.16x10^3$	0.70	-1.03	0.36[b]	-1.41	0.35[b]	3.86±0.19	3.77±0.17	3.29±0.15	3.15±0.02
L145A	0.33	0.69	$1.04x10^3$	0.73	-1.33	0.36[b]	-1.22	0.35[b]	4.01±0.18	3.79±0.02	3.37±0.09	3.26±0.04
D146L	1.03	0.74	$1.45x10^2$	0.80	-1.30	0.36[b]	-0.84	0.35[b]	3.49±0.26	3.22±0.04	2.77±0.22	2.85±0.08
Y131F/D146L	4.53	0.60	$1.41x10^3$	109	-0.98	0.36[b]	-0.59	0.35[b]	4.72±0.14	4.01±0.09	2.69±0.14	2.28±0.14
Δ_{NCap} PP32	6.93	2.57	$4.36x10^3$	2.31	-0.9	0.36[b]	-1.1	0.35[b]	4.13±0.09	4.97±0.10	2.37±0.04	2.65±0.12

Rate constants k in sec^{-1}; $\Delta G^{o}_{H_2O}$ in kcal • mol^{-1}; m values in kcal • mol^{-1} • M^{-1}. [a] For these constructs, parameters were fitted simultaneously to major and minor observed rate constants and unfolding amplitudes. [b] For these variants, which show only one unfolding phase, m_{ID} and m_{NI} were fixed at the wild-type values. Errors for $\Delta G^{o}_{H_2O,eq}$ and m_{eq} are the standard deviation of three independent experiments. Errors for $\Delta G^{o}_{H_2O,kin}$ and m_{kin} (m_{DI}-m_{ID}+m_{IN}-m_{NI}) were determined by propagation assuming independent variables.

A



B

Segment Name	Sequence	Residues
N-CAP	MEMGRRIHLELRNRTPSD	1-18
Repeat 1	VKELVLDNSRSNEGK.....LEG.LTDEFEE	19-43
Repeat 2	LEFLSTINVGLTS.......IAN.LPK.LNK	44-65
Repeat 3	LKKLELSDNRVSGG......LEV.LAEKCPN	66-89
Repeat 4	LTHLNLSGNKIKDLST....IEP.LKK.LEN	90-114
Repeat 5	LKSLDLFNCEVTNLNDYRENVFKLL....PQ	115-141
C-CAP	LTYLDGYDRDDKE	142-154

Figure 3.1. Structure and sequence of PP32. (A) Ribbon representation of the crystal structure of the LRR domain of PP32 (Huyton and Wolberger, 2007). The N-cap, repeats 1-5, and the C-cap are colored as in (B). The residues substituted for Φ-value analysis are shown in CPK representation and colored as in (B). (B) Sequence alignment of the five LRRs of PP32. The five C-terminal residues (150-154) included in this study but not present in the crystal structure are highlighted in yellow.

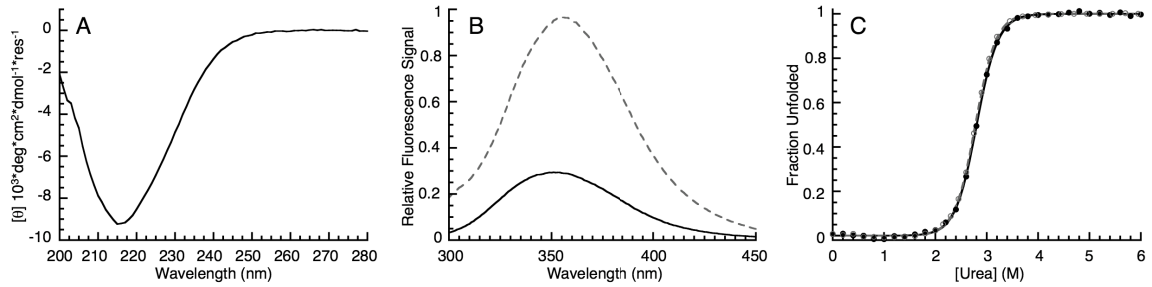


Figure 3.2. Solution spectroscopy and equilibrium unfolding of PP32. (A) Far-UV CD spectrum, showing characteristic β-strand signal with a minimum at 217 nm. (B) Tryptophan fluorescence emission spectra in buffer (continuous line) and 6 M urea (broken line). (C) Urea-induced denaturation monitored by CD at 220 nm (filled circle, continuous line) and fluorescence (open circle, broken line). Lines result from fitting a two-state unfolding model to the data. The CD data are adapted from Dao et al. (2014).

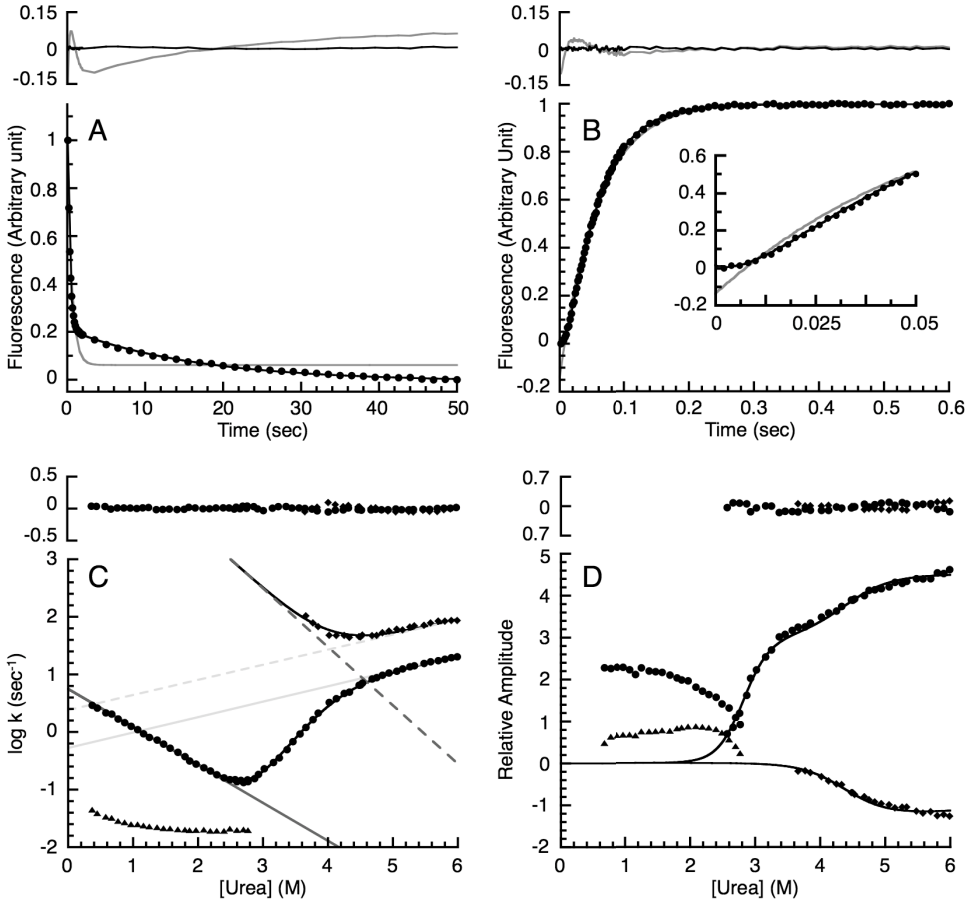


Figure 3.3. Refolding and unfolding kinetics of PP32. (A) Fluorescence-detected refolding (circles) from 4.6 to 0.42 M urea and (B) unfolding to 6 M urea. Gray and black lines show single- and double-exponential fits, respectively; upper panels show residuals. (C) Urea dependence of rate constants and (D) associated amplitudes for folding. Circles, major refolding and unfolding phases; triangles, minor refolding phase; diamonds, minor unfolding phase. The black lines result from fitting of (C) rate constants (excluding the minor proline-limited refolding phase) and (D) unfolding amplitudes using a sequential three-state model (D⇌I⇌N). Straight lines in (C) show fitted folding (dark gray) and unfolding (light gray) rate constants for conversion between D and I (continuous) and between I and N (dashed). Upper panels in (C) and (D) show residuals from the three-state model.

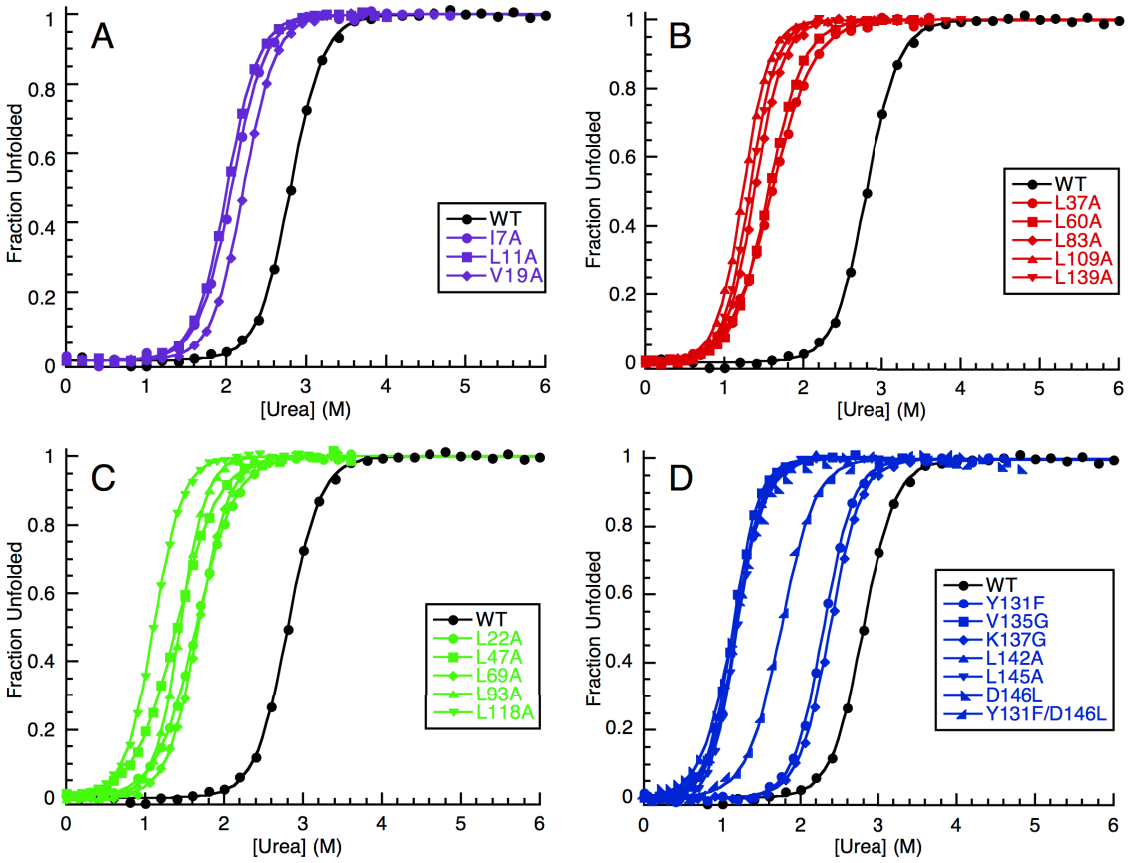


Figure 3.4. Urea-induced equilibrium unfolding of PP32 variants. Transitions are monitored by far-UV CD at 220 nm. Lines result from fitting an equilibrium two-state unfolding model to the data. In each panel, WT PP32 is shown in black. (A) N-cap variants (purple), (B) conserved convex side L→A variants (red), (C) conserved β-sheet L→A variants (green), and (D) C-cap variants (blue).

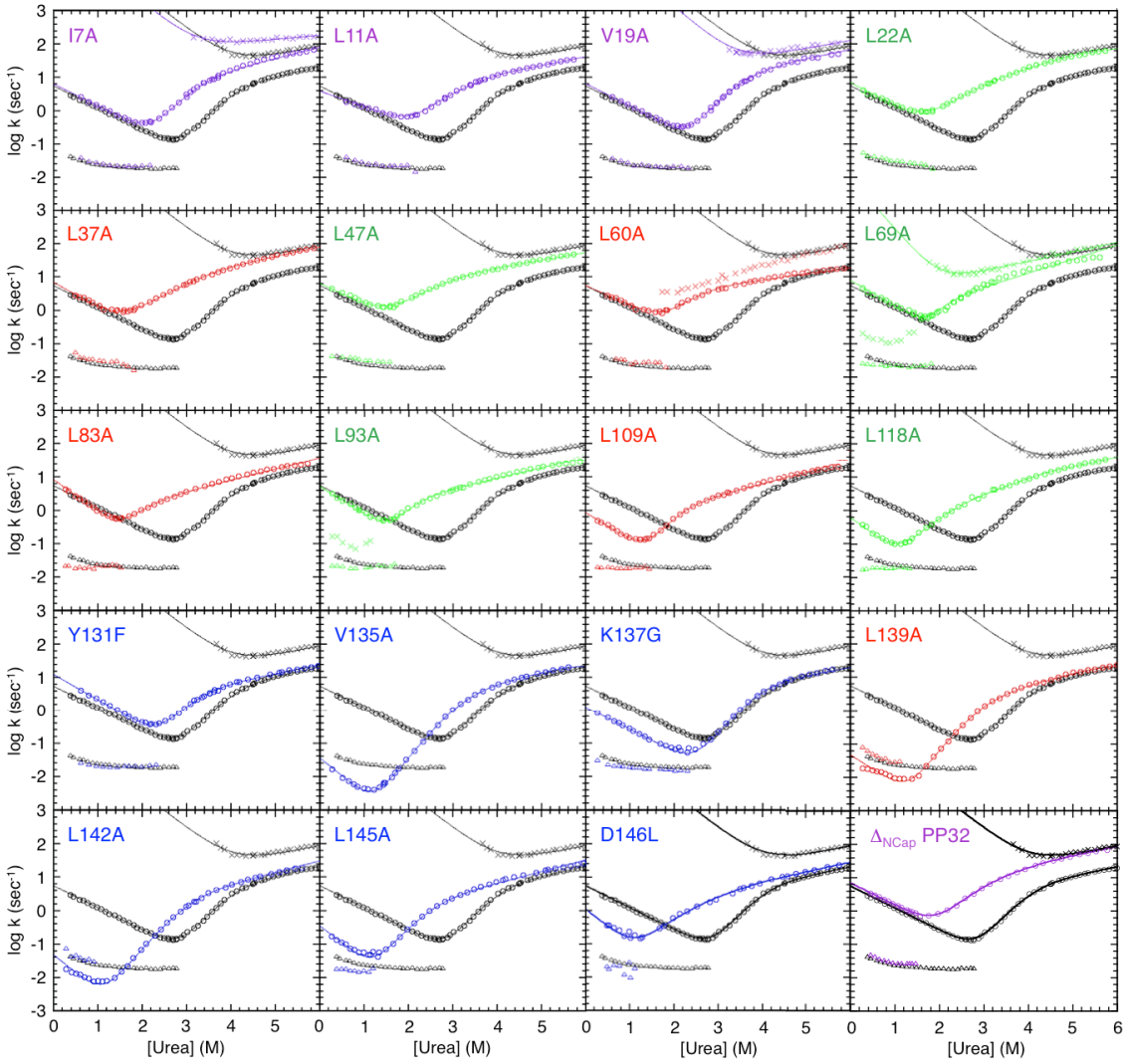


Figure 3.5. Chevron plots for variants of PP32. Urea dependence of fluorescence-monitored rate constants for the major refolding and unfolding phases (circles), the minor refolding phase (triangles), and the minor unfolding phase (diamonds) of WT PP32 (black chevron in each panel), and variants (colors are as in Figure 3.4). Lines result from fitting a sequential three-state kinetic model (D⇌I⇌N) to the data.

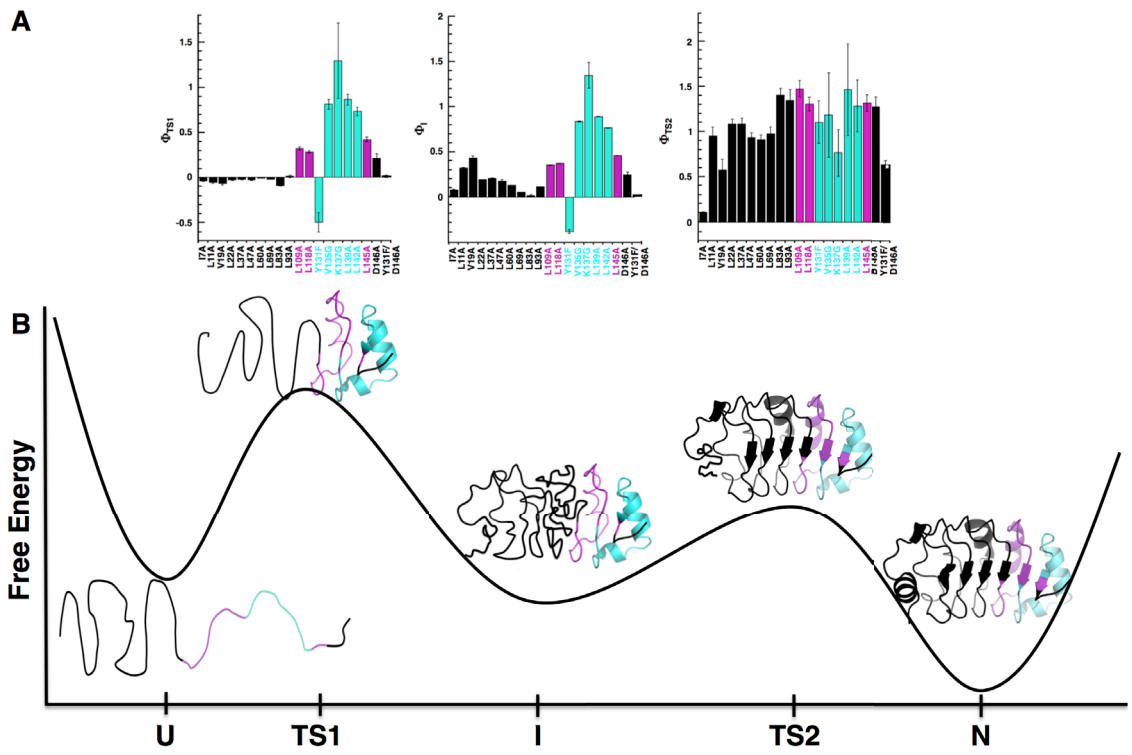


Figure 3.6. Φ-values and the folding pathway of PP32. (A) Folding Φ-values for the first (rate-limiting) transition state (TS1), intermediate (I) and second transition state (TS2). Residues that are structured ($\Phi > 0.5$), partially structured ($0.25 < \Phi < 0.5$) and unstructured ($\Phi < 0.25$) in TS1 are in cyan, magenta and black, respectively. (B) The reaction coordinate for folding of PP32. The structural models of TS1, I, and TS2 are derived from the Φ-values in (A).

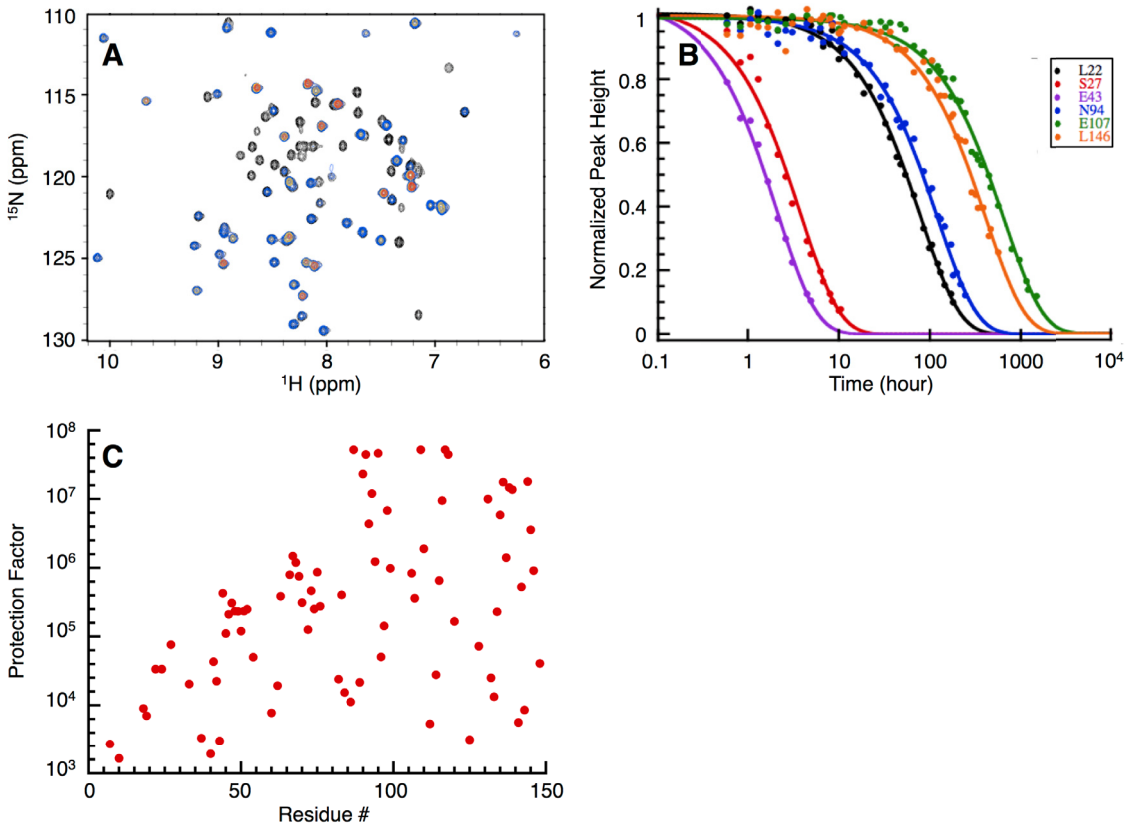


Figure 3.7. Residue-specific hydrogen exchange of PP32. (A) HSQC spectra after PP32 is exchanged from H_2O to D_2O buffer for various times: 21 min (black), 19 hours (blue), 10 days (yellow) and 69 days (red). (B) Exponential decays of peak heights of representative amide protons with very different exchange rates. The lines show single-exponential fits to the data. (C) Protection factors (k_{ex}/k_{int}) for main-chain amides with quantifiable protection. The N-terminus shows uniformly low protection. The region with highest protection factors is toward the C-terminus.

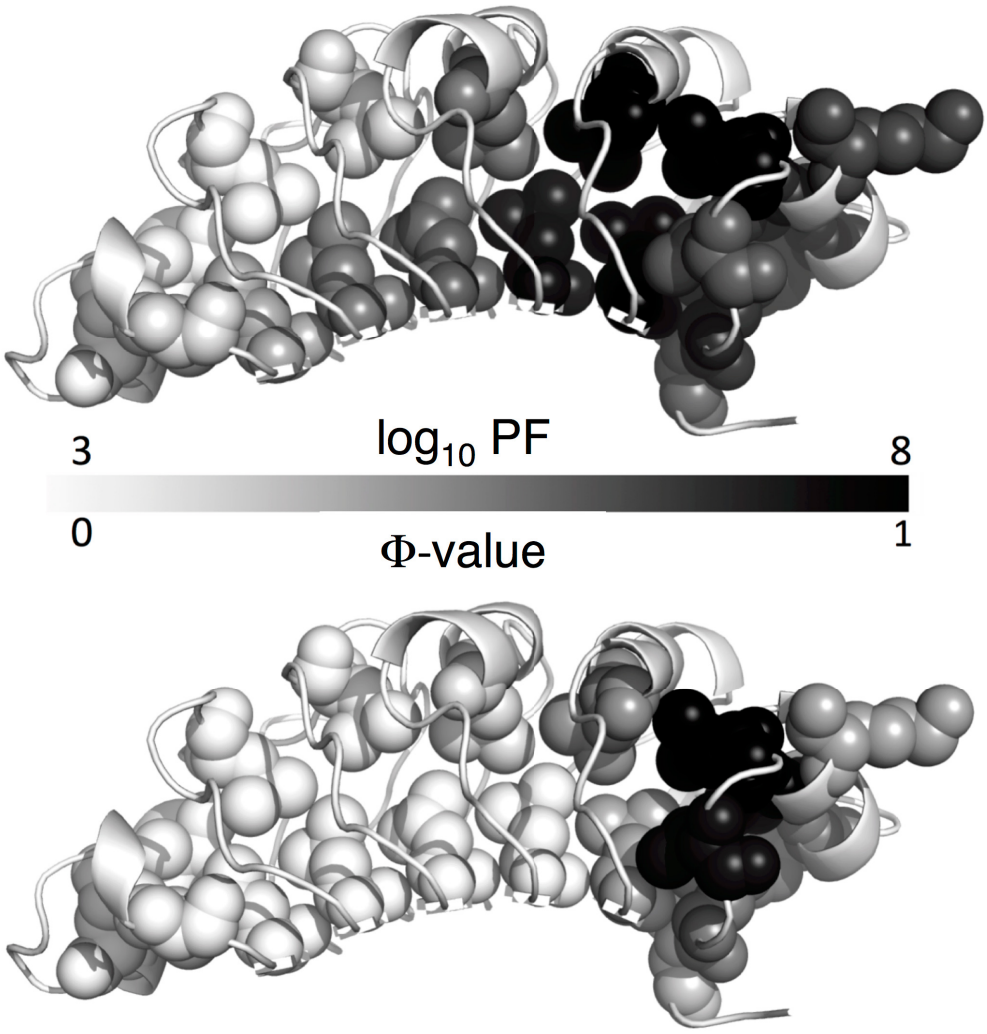


Figure 3.8. Folding pathway of PP32. Ribbon representation, with Φ-value substitutions shown as CPK spheres. Residues are shaded from white to black, with black having the highest local stability (top) and Φ-value (bottom). For direct comparison, only sites with both Φ-value and local stability data are shaded. Based on the coincidence of high protection factors and Φ-values, folding is initiated at the most stable region of PP32.

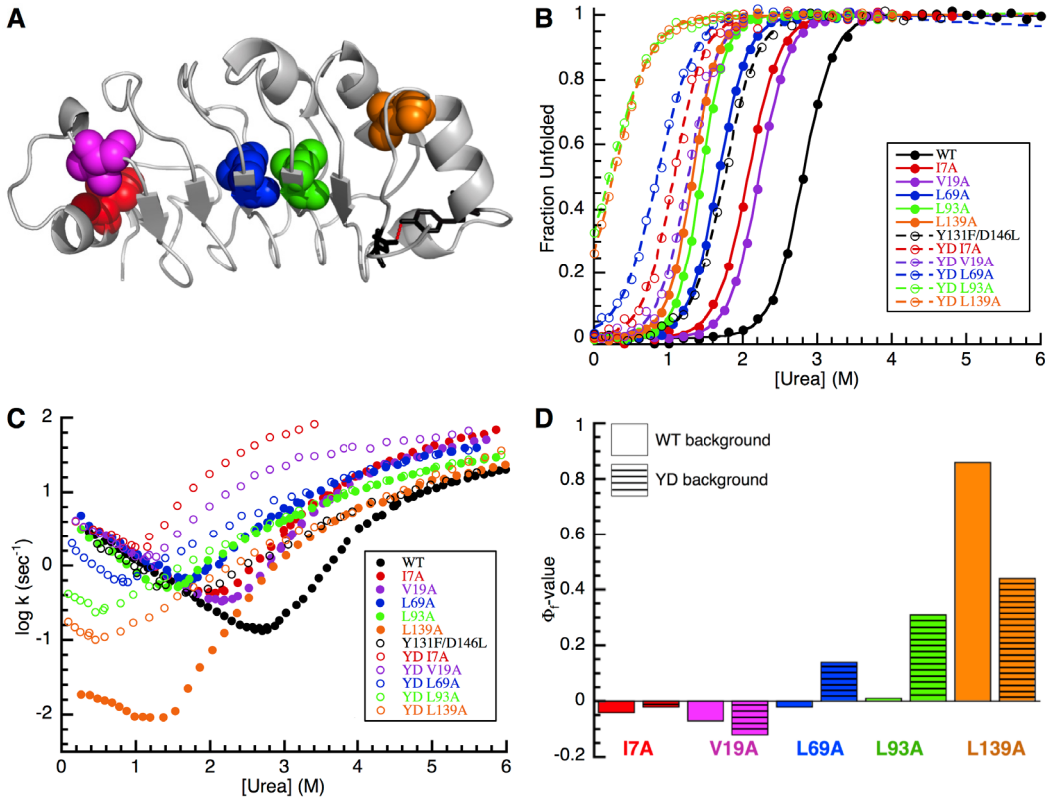


Figure 3.9. Effects of destabilizing the C-terminus to the folding pathway of PP32. (A) Ribbon representation of PP32. Residues Y131 and D146, which we have substituted to destabilize the C-terminus, are shown as black sticks. Residues substituted in Φ-value analysis in the Y131F/D146L (YD) background are shown in CPK representation. (B) Normalized urea-induced unfolding transitions monitored by far-UV CD at 220 nm for variants in WT (closed symbols) and YD (open symbol) backgrounds. Lines result from fitting a two-state equilibrium unfolding model to the data. (C) Urea dependence of fluorescence-monitored rate constants for the major refolding and unfolding phases of variants in WT and YD backgrounds. (D) Φ-values for TS1 in the WT and YD backgrounds. The transition state for the YD construct is less polarized than for WT PP32.

Chapter IV

An Ising analysis of consensus LRR protein folding: high cooperativity from very strong interfacial interactions

4.1 Abstract

To quantify the cooperativity in β-sheet containing leucine-rich repeat (LRR) proteins, and separate intrinsic from interfacial free energies, we have designed a consensus LRR system with protective terminal caps that can be removed one at a time, a necessary condition to resolve intrinsic and interfacial terms. Constructs containing up to eight consensus bacterial leucine-rich repeats (LRR) can be kept soluble and folded in solution when embedded into YopM terminal repeats (N- or C-cap) on either or both ends. The stability and cooperativity of the constructs increase with number of consensus repeats. By globally fitting a nearest-neighbor Ising model to the unfolding transitions of these proteins, we were able to partition the folding free energies into intrinsic interactions within each structural unit (consensus repeats or caps) and interfacial coupling between units. As with α-helical repeat proteins, the intrinsic interactions are destabilizing, whereas the interfacial interactions are stabilizing. However, the consensus LRR system studied here has a significantly greater (more stabilizing) interface than for helical repeat-proteins, making this β-sheet-based repeat system the most cooperative linear repeat protein described to date.

4.2 Introduction

Cooperativity, an essential characteristic of protein folding, is the result of coupled conformational changes between separate structural units to yield either folded or unfolded (but not partially folded) polypeptides. For many globular proteins, regions far apart in sequence can have close contacts and strong coupling, leading to cooperative two-state folding transitions. However, this coupling is difficult to quantify experimentally in globular proteins due to the irregular tertiary structures and sequence-distant contacts that complicate dissection and comparison of different structural units.

In contrast to globular proteins, repeat proteins lack sequence-distant contacts, owing to their elongated and modular architectures. Therefore, repeat proteins might be expected to fold in a noncooperative manner. Surprisingly, two-state equilibrium folding transitions are frequently observed for naturally occurring repeat proteins of different secondary structure motifs, including α-helical repeat proteins (Lowe and Itzhaki, 2007; Tang et al., 1999; Zeeb et al., 2002; Zweifel and Barrick, 2001) and β-strand-containing repeat proteins (Courtemanche and Barrick, 2008; Dao et al., 2014; Kelly et al., 2014; Kloss and Barrick, 2008).

For designed consensus α-helical repeat proteins (ankyrin and tetratricopeptide), a one-dimensional Ising formalism (Zimm and Bragg, 1959) has been used to quantitatively analyze the origin of this cooperativity (Aksel and Barrick, 2009; Aksel et al., 2011; Kajander et al., 2005; Wetzel et al., 2008). Such

consensus proteins, in which each repeat has identical (or nearly identical) sequence, reduce the complexity of folding energetics so that just two energy terms are required to describe the consensus array: an intrinsic folding energy for each repeat, and a pairwise interfacial coupling energy between nearest neighbors.[1]

For both consensus ankyrin repeat and tetratricopeptide repeat (TPR) constructs, the intrinsic folding of each repeat is unfavorable, whereas the interfacial interactions between the repeats are highly favorable. This energy distribution leads to coupled folding between individual repeats. However, TPR proteins show a much lower level of cooperativity, due to both a weaker interfacial coupling and a less unfavorable intrinsic folding energy. The structural origins of these differences are unclear.

Can a similar distribution of intrinsic and interfacial energies be the origin of cooperativity in the folding β-strand-containing repeat proteins? Whereas the average buried surface areas at the interfaces of β-strand repeat and α-helical repeat proteins are similar, the change in surface area upon folding of single β-strand repeat is significantly lower than that of α-helical repeat units (Kloss et al., 2008). Therefore, the intrinsic folding of β-strand repeats might be expected to be even more unfavorable, leading to higher cooperativity in the folding of these proteins. Indeed, for naturally occurring β-strand leucine-rich-repeat (LRR) proteins, *m*-values of unfolding, which provide an estimate of the size of the

[1] Typically, two additional energy terms are needed to capture contributions from the terminal capping sequences, which are modified repeats required to maintain solubility.

cooperative unit, are higher than predicted for the size of the constructs (Courtemanche and Barrick, 2008; Dao et al., 2014; Kelly et al., 2014; Kloss and Barrick, 2008). However, for the two sets of consensus LRR proteins studied to date, based on the RI-subfamily, *m*-values are much lower than expected and do not increase with the length of the constructs (Parker et al., 2014; Stumpp et al., 2003). Together with the low unfolding free energies, the low *m*-values suggest that these consensus LRR proteins fold in a modular, noncooperative manner, or that they are associated in solution. Either way, these two consensus LRR systems are not amenable to thermodynamic analysis.

To gain insights to the cooperativity observed in naturally occurring LRR proteins, we sought to create and study consensus LRR constructs with folding properties similar to their natural counterparts. We chose the bacterial LRR protein subfamily, since the folding of one of its members, YopM, has been studied extensively and has been shown to have very high level of cooperativity (Evdokimov et al., 2001; Kloss and Barrick, 2008, 2009; Vieux and Barrick, 2011). Moreover, bacterial LRRs are very regular in sequence, secondary structure, and repeat length, making alignment straightforward. Here, we show that these consensus LRR constructs fold cooperatively and are amenable to Ising analysis.

4.3 Results

Consensus design

Here we describe an iterative approach that lead us to the soluble, folded, and stable consensus arrays used in the thermodynamic analysis below. This approach involved both refinement of the sequences within internal repeats, and inclusion of several different sequence caps LRR proteins are often observed with capping motifs that are essential for structure formation of the repeats (Breitsprecher et al., 2014; Dao et al., 2014). For the bacterial LRR proteins, the N-terminus is anchored by an α-helical cap (Evdokimov et al., 2001; Keszei et al., 2014). Therefore, consensus repeats and cap might be necessary for well-behaved, stable constructs.

Alignment of bacterial LRR protein repeats can be used to generate a hidden Markov model (HMM, Punta et al., 2012) representation of the subfamily, which can be displayed graphically as a sequence logo (Figure 4.1A). In the first round of design, we picked the most frequently occurring amino acid at each position (tallest letter in the logo) to create consensus “R” repeats (Figure 4.1B), an approach used successfully for consensus α-helical repeat proteins. In contrast to these helical consensus proteins, polypeptides containing up to 10 R’s, with or without the designed N-terminal α-helical cap, were unfolded in solution (data not shown). The inability of these first-round constructs to fold may result from the high negative charge of each consensus sequence, resulting in unfavorable interactions between the repeats.

For several positions in R, the most frequently occurring residue is only marginally more preferred than some other amino acids at that same position (Figure 4.1A). Moreover, inspection of sequences of pairs of LRRs reveals that positive and negative charges often alternate at the same position of adjacent repeats. For example, at position one of YopM repeats, an E in one repeat is often followed by a K in the next. To achieve a more balanced charge distribution, we made substitutions to these alternating positions to create consensus repeats A and B (residues highlighted in magenta, Figure 4.1B). The side chains of the residues substituted for charge balance are solvent-exposed (Figure 4.1C). The remaining positions, including the highly conserved core residues (highlighted in green, Figures 4.1B and 4.1C), are identical from repeat to repeat. However, constructs containing multiple "AB" units were insoluble, even with a consensus N-terminal cap (not shown). This insolubility could be caused by exposure of one or both ends of the AB consensus array to solvent. Such association has been problematic in studies of helical repeat proteins, where solubilizing capping repeats have been required to enhance solubility.

To eliminate this potential solvent exposure, we embedded AB repeat pairs internally within YopM terminal repeats (N-3 and 13-15.5, Figure 4.1D). In contrast to the uncapped and consensus-capped AB repeats, these YopM-embedded AB repeats are soluble at high concentrations. These YopM terminal sequences will be referred to as "N" and "C" sequences, respectively. For this work, we will study constructs with either or both N and C caps, with varying

number of alternating A's and B's (Figure 4.1E). Velocity sedimentation analytical ultracentrifugation (VS AUC) studies show that representative construct capped by N only, $N(AB)_2A$, is monomeric at all concentrations studied (up to 51 μM, Figure 4.2A). The representative doubly capped construct $N(AB)_3C$ (three AB units flanked by N and C) is also monomeric in solution by VS AUC at low protein concentrations, where urea-induced denaturation is carried out (Figure 4.2B). Urea-induced unfolding transitions are independent of protein concentrations (Figure 4.2C). The single-cap constructs are critical for nearest-neighbor analysis, allowing the stability contributions of the caps to be separated from the internal consensus repeats.

Solution properties of consensus bacterial LRR proteins

All N- and C-capped consensus AB LRR constructs are folded, based on far-UV CD spectroscopy (Figure 4.3A). The far-UV CD spectra have different shapes, consistent with the constructs having different amounts of α-helix and β-sheet, depending on the cap composition and length of the AB array. Constructs without N do not contain helical structure, and therefore, have characteristic β-strand far-UV CD spectra. In contrast, spectra of constructs with N show less helical characteristic as more β-stranded repeats are added.

Urea-induced equilibrium unfolding of consensus bacterial LRR proteins

To compare the stabilities of the different consensus constructs, we monitored urea-induced unfolding transitions using far-UV CD spectroscopy

(Figure 4.3B). Both the midpoint and steepness of the transitions increase with the number of consensus repeats, as observed for consensus helical repeats. Analysis of the unfolding transitions separately using a two-state model confirms that the denaturant sensitivities of folding of all constructs increase with size (Table 4.1).

The AB LRR consensus proteins are significantly more stable than the naturally occurring YopM. The most direct comparison is between $N(AB)_2C$ and YopM, both of which have the same N- and C-termini (100 and 66 residues, respectively), but the middle nine repeats of YopM (LRRs 4-12) are replaced by four consensus repeats (two AB units). Despite being significantly smaller in size, $N(AB)_2C$ has very similar stability compared to YopM (-12.9 kcal • mol^{-1} in 500 mM NaCl vs. -12.4 kcal • mol^{-1} in 412 mM NaCl and -13.1 kcal • mol^{-1} in 578 mM NaCl; Kloss and Barrick, 2008). Given the trend we see for stability to increase with AB units, we expect that a consensus construct of comparable size to YopM [$N(AB)_4C$] would be considerably more stable.

Aside from keeping the constructs soluble, the caps and their interfacial couplings with the consensus repeats also contribute significantly to the folding stability of these constructs. For example, doubly capped construct $N(AB)_3C$, with an $\Delta G^{o}_{H_2O}$ of 16.5 kcal • mol^{-1}, is much more stable than singly capped constructs $(AB)_3C$ and $N(AB)_3$, with $\Delta G^{o}_{H_2O}$ of 1.9 and 7.2 kcal • mol^{-1}, respectively. Thus, the N-cap contributes more to the stability than the C-cap.

One-dimensional Ising analysis of the urea-induced unfolding transitions

To determine the distribution of intrinsic folding and interfacial energies for this consensus LRR series, we globally fitted an Ising model to the unfolding transitions (Figure 4.4). This model differs from the equilibrium two-state model in that it includes all of the possible partially folded conformations, in addition to the fully folded and fully unfolded conformations, in the folding transition. Each repeat is treated as an individual folding unit that undergoes conformational transition independent of the non-neighboring repeats. The model includes intrinsic folding energies for N, consensus AB unit, and C (ΔG_N, ΔG_{AB}, and ΔG_C). The interfacial energy, $\Delta G_{i,i+1}$, is assumed to be the same for the interfaces between N-cap and A, B and A, and B and C-cap. The free energy of each repeat is linearly dependent on urea concentration, with the denaturant dependence, m_i, assumed to be the same for the N-cap, consensus AB, and the C-cap.

The Ising model, with five globally shared thermodynamic parameters discussed above, fits well to the seven well-resolved unfolding transitions from different constructs (Table 4.2). The intrinsic folding of individual AB structural units is highly unfavorable at 10.5 kcal/mol. Folding of the N unit is the least unfavorable at 2.5 kcal/mol, a likely reflection of its larger size. The disfavored intrinsic energies are offset by the highly favorable interfacial coupling between the units, with a free energy of -14.1 kcal/mol per interface. This partitioning of energies is consistent with the high level of cooperativity observed for these constructs: single folded units are energetically unfavorable, but when two or

more (depending on which cap is present) adjacent units are folded, the interfacial coupling between them drives the reaction towards fully folded state.

4.4 Discussion

The motivation for this study is to gain insights into the origin of cooperativity observed for the folding of several naturally occurring β-strand containing LRR proteins (Courtemanche and Barrick, 2008; Dao et al., 2014; Kelly et al., 2014; Kloss and Barrick, 2008), and to better understand the structural and thermodynamic origins of cooperativity in general. The simple linear, modular architecture of repeat proteins has permitted the use of one-dimensional nearest-neighbor Ising models in describing the stability and cooperativity of α-helical repeat proteins (Aksel and Barrick, 2009; Aksel et al., 2011; Kajander et al., 2005; Mello and Barrick, 2004; Wetzel et al., 2008). With the exception of the work by Mello and Barrick, the above studies extend the structural redundancy of α-helical repeat proteins to the level of primary structure, using identical (or nearly identical) consensus repeats. This simplification enables intrinsic folding and interfacial interaction energies to be resolved unambiguously.

To date, efforts to create stable consensus LRR proteins that fold cooperatively have been unsuccessful. The ability to create consensus LRR constructs with the high cooperativity observed for naturally occurring LRR proteins will provide insights to the folding of not only this family of repeat proteins, but also globular proteins.

Designed consensus LRR proteins are stable and undergo cooperative unfolding transitions

After several rounds of designing, we were able to obtain well-behaved constructs containing consensus bacterial LRRs. Obtaining folded consensus-based LRR constructs required the alternating A and B consensus repeats that contain oppositely charged residues at surface positions, as is often observed naturally occurring bacterial LRR proteins. Maintaining solubility required the addition of the terminal repeats of YopM at one or both ends of the consensus repeats. Capping motifs have been shown to be essential for the folding of YopM (Kloss and Barrick, 2009). However, the interactions between caps and consensus repeats are likely to be highly specific since the constructs aggregate even in the presence of a designed α-helical N-terminal cap.

Denaturation curves for all constructs are sigmoidal and can be well fitted by an equilibrium two-state model (Figure 4.3B). Both free energies of unfolding and *m*-values increase with the number of consensus repeats added. The *m*-values are higher than predicted based on the constructs lengths, as observed for naturally occurring LRR proteins that fold by equilibrium two-state mechanisms (Courtemanche and Barrick, 2008; Dao et al., 2014; Kelly et al., 2014; Kloss and Barrick, 2008). For the two previously reported sets of consensus LRR proteins, *m*-values are much lower than predicted and do not increase with number of repeats (Parker et al., 2014; Stumpp et al., 2003). The stabilities of the consensus constructs are much higher for YopM constructs of the same length. This increased stability in consensus LRR constructs has also

been observed for consensus α-helical repeat proteins (Aksel et al., 2011; Tripp and Barrick, 2007; Wetzel et al., 2008).

The partition of free energies is consistent with high cooperativity in folding

The one-dimensional nearest-neighbor Ising formalism provides a means for the thermodynamics to be described using a limited set of parameters. For set of consensus LRR constructs described here, five parameters, including one for interfacial interactions between the structural units, three for intrinsic folding of each unit (N, AB, and C), and one for denaturant dependence, can be determined uniquely. This Ising model can be well-fitted to the data (Figure 4.4), yielding well-determined parameters (Table 4.2).

Cooperativity can be observed if intrinsic folding is unfavorable such that individual units cannot stay folded on their own, but interfacial interactions are stabilizing enough to drive folding of the entire construct. Indeed, for the consensus LRR constructs studied here, all three intrinsic folding energies are unfavorable, whereas the interfacial interactions are highly favorable. The intrinsic folding of an AB consensus unit, at +10.5 kcal/mol, is much more destabilizing than the intrinsic folding of all three consensus α-helical repeats studied to date, at +2.3 to +5.4 kcal/mol (Aksel et al., 2011; Kajander et al., 2005; Wetzel et al., 2008). In contrast, the interfacial interactions for the LRR constructs, at -14.1 kcal/mol per interface, is more stabilizing than those of the consensus ankyrin repeat constructs, at about -12 kcal/mol per interface, and much more stabilizing for the consensus TPR constructs at -4.5 kcal/mol per repeat. This

extreme mismatch of the intrinsic and interfacial energies for the consensus LRR constructs, compared to the α-helical consensus constructs, results in the very high level of cooperativity observed for these proteins as well as their naturally occurring counterparts.

Though the present analysis treats the AB pair as a single unit, each pair contains two individual repeats and one A/B interface. Though we cannot resolve any differences in stability between the A and B repeats, in the simplest scenario, in which A and B have the same intrinsic folding energy and A/B interface is the same B/A interface, we would have:

$$10.5 = \Delta G_A + \Delta G_B + \Delta G_{A/B} = 2\Delta G_A + \Delta G_{B/A} \quad (4.1)$$

or

$$\Delta G_A = \frac{10.5 - \Delta G_{B/A}}{2} \quad (4.2)$$

The term $\Delta G_{B/A}$ in Equation 4.2 is simply the coupling energy in the Ising model, and can be determined directly from our fit of the unfolding transitions (-14.1 kcal/mol, Table 4.1). Substituting this value into Equation 4.2 gives an intrinsic folding energy of about +12.3 kcal/mol, which is extremely destabilizing compared to intrinsic folding energies of α-helical repeats.

The following structural features may contribute to the differences in the intrinsic folding and interfacial interactions between two types of repeats: 1) the change in surface area upon folding of a LRR repeat is about half of that of a

helical repeat, 2) peptide hydrogen bonds form during the intrinsic folding of helical repeat units, though they are not formed until interface formation in LRRs 3) the helix-helix interaction within individual TPR and ankyrin repeats may lead to intrinsic stability. These observations are consistent with significantly more unfavorable folding of individual LRRs compared to α-helical repeats.

The structural origins of the difference in interfacial coupling between LRR and helical repeats are harder to explain, especially since the difference in the interfacial energies between the TPRs and ankyrin repeats is more pronounced than that between the ankyrin repeats and LRRs. The types of interactions between LRRs are fundamentally different from those between helical repeats: adjacent LRRs interact in part by formation of β-sheet hydrogen bonds, whereas for TPRs and ankyrin repeats, interactions are mostly helix packing. Within LRR proteins, the network of interactions between the repeats is extensive, including packing of the hydrophobic cores of the repeats, backbone hydrogen bonds between the β-strands, electrostatic interactions between the alternating charged residues, and hydrogen bonds between the side chain NH hydrogens of conserved asns of the asparagine ladder (Figure 4.1D) to backbone carbonyl oxygens of preceding repeats. These interactions may all contribute to favorable interface formation between LRRs.

4.5 Conclusions

Repeat proteins contain different types of secondary structural elements and interactions, including burial and packing on hydrophobic side chains,

hydrogen bonds, as well as charge-charge interactions, all of which are found in globular proteins. Therefore, investigations into the folding of repeat proteins provide direct insights to the folding of globular proteins. Our work here on consensus β-strand-containing LRR proteins expands previous knowledge gained from studies of α-helical repeat proteins: cooperativity in folding is the result of favorable interactions between structural units to balance out the unfavorable interactions within structural units, although the degree of cooperativity highly depends on the type of secondary structures involved. Additional studies of other repeats, containing different types and compositions of secondary structures, as well as point substitutions within our LRR framework, will provide better generalizations to globular protein folding.

4.6 Materials and Methods

Subcloning, protein expression, and purification

Point substitutions were made using Quikchange (Stratagene, La Jolla, CA) to gene encoding YopM (Kloss and Barrick, 2008) contained in pET15-b expression vector such that BamHI and BglII sites were at the end of LRR3 and LRR12, respectively. For doubly capped consensus constructs, oligonucleotides (IDT; Coralville, CA) for genes encoding consensus LRRs A and B were phosphorylated and ligated into modified vector between the BamHI and BglII sites. The length of the consensus were propagated as described for consensus ankyrin (Aksel et al., 2011). To create constructs capped only with N, point mutations to change the end of doubly capped constructs, before the Histag, to a BglII site. These modified vectors were next digested with BglII, purified and ligated. For constructs capped only with C, no expression was observed. When an expression tag (LRR protein PP32, Dao et al., 2014) was added, separated by a TEV protease site, high expression was observed.

Constructs were expressed in Escherichia coli BL21 cells in Luria broth at 20^{o}C overnight. Bacteria were pelleted, lysed in 20mM $NaPO_4$, 500mM NaCl, 25mM imidazole, 0.1mM TCEP (pH 7.4), and purified via Ni^{2+} chromatography. For constructs without N cap, purified proteins were dialyzed into 50mM TrisHCl, 0.5 mM EDTA, 1 mM DTT, 1 M urea (the presence of urea helped with digest), pH 8, then digested with TEV protease and purified via an anion exchange column. Purified proteins were dialyzed into 20 mM Na PO_4, 500 mM NaCl, 0.1

mM TCEP (pH 7.8), and were frozen at -80°C. Protein concentrations were determined as described (Edelhoch, 1967).

Analytical ultracentrifugation (AUC) sedimentation velocity (SV)

Experiments were performed using a ProteomeLab Beckman XL-I analytical ultracentrifuge. Prior to AUC experiments, all proteins were extensively dialyzed into 20mM Sodium Phosphate, 500mM NaCl, 0.1mM TCEP, 0.1mM EDTA, pH 7.8, and the resulting dialysate saved for reference. Protein samples were prepared to span a wide concentration range (~5-100μM) by dilution into dialysate buffer. Solvent density and viscosity were calculated using the program Sednterp (Tom Laue, John Philo, Biomolecular Interaction Technologies Center).

AUC SV cells were assembled using SedVel60K 1.2mm meniscus-matching centerpieces (SpinAnalytical) and sapphire windows. All other cell components were purchased from Beckman Coulter. Upon loading sample and reference, the cells were properly aligned in an An60Ti 4-hole rotor and spun @3-6krpm for approximately 5-10 minutes until the menisci of the sample and reference matched precisely. Slightly more reference buffer was added to the reference sector to prevent sample flow into the reference sector. As a result, protein sample concentrations were slightly diluted. After meniscus matching, the rotor and cells were removed from the AUC, and inverted multiple times to remix samples. The rotor containing the aligned cells was then reinserted into the AUC chamber, and allowed to temperature equilibrate under vacuum @20°C for at least 90 minutes. SV experiments were run for about 8 hours @50krpm.

Circular Dichroism spectroscopy

All CD measurements were carried out using Aviv Model 400 CD spectrometer (Lakewood, NJ). Far-UV CD spectra were collected in an 0.1 cm path-length quartz cuvette with protein concentrations ranging from 15 to 30 μM. Spectra were averages of three wavelength scans, each with 1 nm step size and 10 seconds signal average per step.

Equilibrium unfolding

Urea-induced unfolding was monitored by CD at 220 nm. Urea (Amresco, Solon, OH) was deionized by chromatography over mixed-bed resin (Bio-Rad, Hercules, CA). Urea concentration was determined by refractometry.(Pace, Nick C, 1986) Urea titrations were carried out using a computer-controlled Microlab syringe titrator (Hamilton, Reno, NV). Samples contained 2-4 μM protein, 20 mM $NaPO_4$, 500 mM NaCl, 0.1 mM TCEP (pH 7.8). At each urea concentration, samples were equilibrated for 5-10 minutes at 20^{o}C and CD signal was averaged for 30 seconds. Two-state analysis of equilibrium unfolding transitions were carried out as described by Street et al (2008).

To determine the interfacial and intrinsic energetic terms, unfolding transitions of the consensus LRR constructs were globally fitted with a heteropolymer nearest-neighbor Ising model using in-house scripts.

4.7 References

Aksel, T., and Barrick, D. (2009). Chapter 4 Analysis of Repeat-Protein Folding Using Nearest-Neighbor Statistical Mechanical Models. In Methods in Enzymology, J.M.H. and G.K.A. Michael L. Johnson, ed. (Academic Press), pp. 95–125.

Aksel, T., Majumdar, A., and Barrick, D. (2011). The Contribution of Entropy, Enthalpy, and Hydrophobic Desolvation to Cooperativity in Repeat-Protein Folding. Structure *19*, 349–360.

Breitsprecher, D., Gherardi, E., Bleymuller, W.M., and Niemann, H.H. (2014). Crystal structure of an engineered YopM-InlB hybrid protein. BMC Struct. Biol. *14*, 12.

Courtemanche, N., and Barrick, D. (2008). Folding thermodynamics and kinetics of the leucine-rich repeat domain of the virulence factor Internalin B. Protein Sci. Publ. Protein Soc. *17,* 43–53.

Dao, T.P., Majumdar, A., and Barrick, D. (2014). Capping motifs stabilize the LRR protein PP32 and rigidify adjacent repeats. Protein Sci. Publ. Protein Soc.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry (Mosc.) *6*, 1948–1954.

Evdokimov, A.G., Anderson, D.E., Routzahn, K.M., and Waugh, D.S. (2001). Unusual molecular architecture of the Yersinia pestis cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit. J. Mol. Biol. *312*, 807–821.

Kajander, T., Cortajarena, A.L., Main, E.R.G., Mochrie, S.G.J., and Regan, L. (2005). A New Folding Paradigm for Repeat Proteins. J. Am. Chem. Soc. *127*, 10188–10190.

Kelly, S.E., Meisl, G., Rowling, P.J.E., McLaughlin, S.H., Knowles, T., and Itzhaki, L.S. (2014). Diffuse transition state structure for the unfolding of a leucine-rich repeat protein. Phys. Chem. Chem. Phys.

Keszei, A.F.A., Tang, X., McCormick, C., Zeqiraj, E., Rohde, J.R., Tyers, M., and Sicheri, F. (2014). Structure of an SspH1-PKN1 Complex Reveals the Basis for Host Substrate Recognition and Mechanism of Activation for a Bacterial E3 Ubiquitin Ligase. Mol. Cell. Biol. *34*, 362–373.

Kloss, E., and Barrick, D. (2008). Thermodynamics, Kinetics, and Salt dependence of Folding of YopM, a Large Leucine-rich Repeat Protein. J. Mol. Biol. *383*, 1195–1209.

Kloss, E., and Barrick, D. (2009). C-terminal deletion of leucine-rich repeats from YopM reveals a heterogeneous distribution of stability in a cooperatively folded protein. Protein Sci. Publ. Protein Soc. *18*, 1948–1960.

Kloss, E., Courtemanche, N., and Barrick, D. (2008). Repeat-protein folding: New insights into origins of cooperativity, stability, and topology. Arch. Biochem. Biophys. *469*, 83–99.

Lowe, A.R., and Itzhaki, L.S. (2007). Biophysical Characterisation of the Small Ankyrin Repeat Protein Myotrophin. J. Mol. Biol. *365*, 1245–1255.

Mello, C.C., and Barrick, D. (2004). An experimentally determined protein folding energy landscape. Proc. Natl. Acad. Sci. U. S. A. *101*, 14102–14107.

Pace, Nick C (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzym. *131*, 266–280.

Parker, R., Mercedes-Camacho, A., and Grove, T.Z. (2014). Consensus design of a NOD receptor leucine rich repeat domain with binding affinity for a muramyl dipeptide, a bacterial cell wall fragment. Protein Sci. n/a–n/a.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40*, D290–D301.

Street, T.O., Courtemanche, N., and Barrick, D. (2008). Protein Folding and Stability Using Denaturants. In Methods in Cell Biology, I. Dr. John J. Correia and Dr. H. William Detrich, ed. (Academic Press), pp. 295–325.

Stumpp, M.T., Forrer, P., Binz, H.K., and Plückthun, A. (2003). Designing Repeat Proteins: Modular Leucine-rich Repeat Protein Libraries Based on the Mammalian Ribonuclease Inhibitor Family. J. Mol. Biol. *332*, 471–487.

Tang, K.S., Guralnick, B.J., Wang, W.K., Fersht, A.R., and Itzhaki, L.S. (1999). Stability and folding of the tumour suppressor protein p16. J. Mol. Biol. *285*, 1869–1886.

Tripp, K.W., and Barrick, D. (2007). Enhancing the Stability and Folding Rate of a Repeat Protein through the Addition of Consensus Repeats. J. Mol. Biol. *365*, 1187–1200.

Vieux, E.F., and Barrick, D. (2011). Deletion of internal structured repeats increases the stability of a leucine-rich repeat protein, YopM. Biophys. Chem. *159*, 152–161.

Wetzel, S.K., Settanni, G., Kenig, M., Binz, H.K., and Plückthun, A. (2008). Folding and Unfolding Mechanism of Highly Stable Full-Consensus Ankyrin Repeat Proteins. J. Mol. Biol. *376*, 241–257.

Zeeb, M., Rösner, H., Zeslawski, W., Canet, D., Holak, T.A., and Balbach, J. (2002). Protein folding and stability of human CDK inhibitor p19INK4d. J. Mol. Biol. *315*, 447–457.

Zimm, B.H., and Bragg, J.K. (1959). Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains. J. Chem. Phys. *31*, 526–535.

Zweifel, M.E., and Barrick, D. (2001). Studies of the Ankyrin Repeats of the Drosophila melanogaster Notch Receptor. 2. Solution Stability and Cooperativity of Unfolding†. Biochemistry (Mosc.) *40*, 14357–14367.

Table 4.1. Thermodynamic unfolding parameters different of consensus LRR constructs

	$\Delta G^{o}_{H_2O}$	*m-value*	C_M
N(AB)	-1.10 ± 0.08	1.53 ± 0.04	0.72 ± 0.05
$N(AB)_2$	-4.48 ± 0.11	2.53 ± 0.11	1.77 ± 0.09
$N(AB)_3$	-7.00 ± 0.38	2.98 ± 0.16	2.35 ± 0.01
$(AB)_3C$	-1.77	2.58	0.69
$(AB)_4C$	-5.10	3.38	1.51
N(AB)C	-9.54 ± 0.19	3.37 ± 0.04	2.83 ± 0.02
$N(AB)_2C$	-12.87 ± 0.27	4.12 ± 0.08	3.12 ± 0.01
$N(AB)_3C$	-16.45 ± 0.39	5.09 ± 0.11	3.23 ± 0.02

Values determined from urea-induced denaturation. $\Delta G^{o}_{H_2O}$ in kcal • mol^{-1}; *m* values in kcal • mol^{-1} • M^{-1}, C_M in M (urea).

Table 4.2. Parameters obtained from fitting of the Ising model to unfolding transitions of consensus LRR constructs

ΔG_N	ΔG_{AB}	ΔG_C	$\Delta G_{i,i+1}$	m_i
2.5	10.5	7.7	-14.1	0.9

Free energy is in kcal • mol^{-1}; *m*-value is in kcal • mol^{-1} • M^{-1}. The intrinsic energy for folding of a structural unit, N- or C-cap, or consensus AB, is represented as ΔG_i. The interfacial energy between N and A, B and A, or B and C is represented as $\Delta G_{i,i+1}$. The denaturant dependence of a structural unit is represented as m_i.

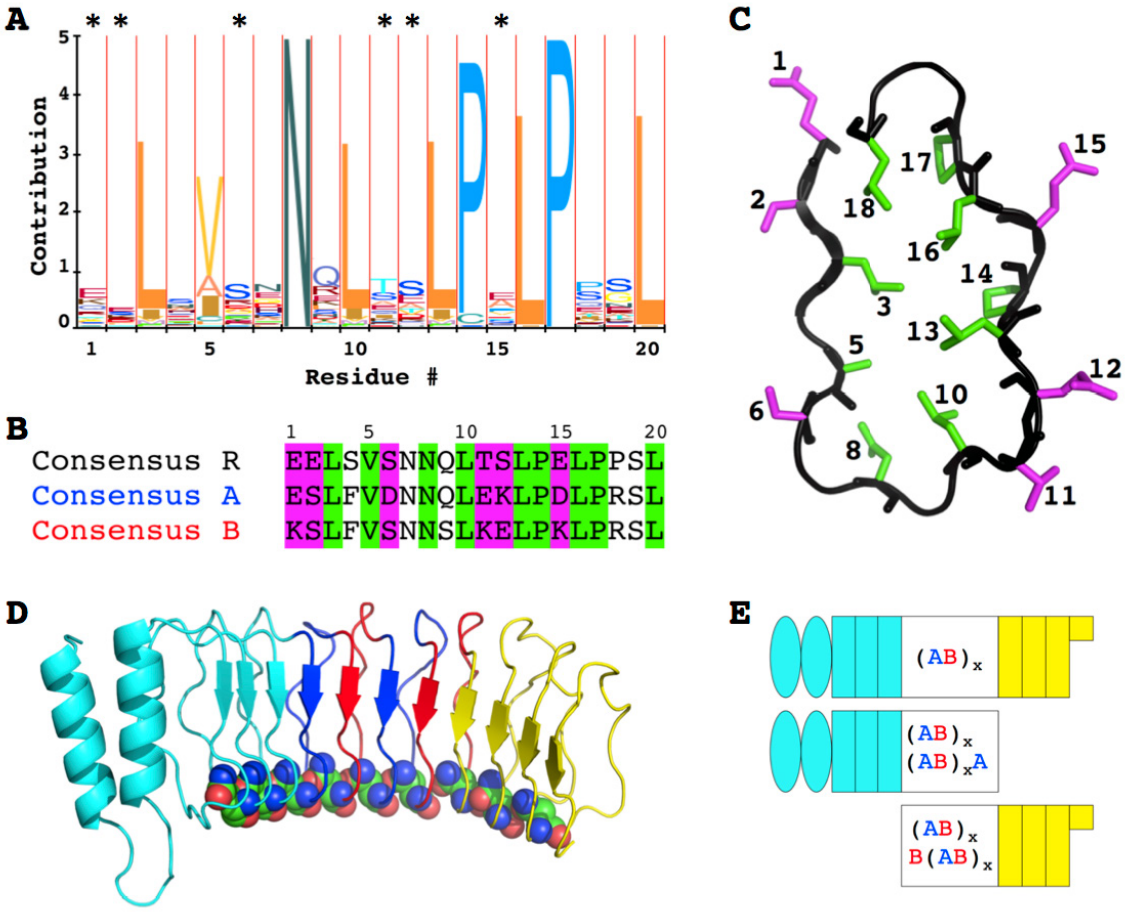


Figure 4.1. Design of consensus bacterial LRR protein repeats. (A) Hidden Markov model (HMM) representation of the bacterial LRR protein subfamily. The height of each letter corresponds to the frequency at which an amino acid occurs at a specific position. Positions where charges were varied between A and B repeats are indicated with asterisks. (B) Sequences of consensus R repeat obtained from an HMM, and of modified consensus A and B used in this study. Highly conserved residues are highlighted in green. Variable residues substituted for charge balance are highlighted in magenta. (C) Ribbon diagram of a representative bacterial leucine-rich repeat, LRR2 of YopM (residues 94-113, PDB: 1JL5, Evdokimov et al., 2001). The residues highlighted in (B) are shown in

sticks in the same colors. All of the conserved residues point towards the core of the protein. The variable residues are solvent exposed. (D) Model of a designed construct, $N(AB)_2C$ based on the YopM crystal structure (PDB: 1JL5, Evdokimov et al., 2001). Capping motif “N” (YopM N-terminal cap and LRRs 1-3) is shown in cyan. Capping motif “C” (LRRs 13-15.5) is shown in yellow. Consensus repeats “A” and “B” are shown in blue and red, respectively. The conserved asparagine residue at position eight of each repeat is shown in spheres. (E) Schematic representations of the different consensus constructs with both capping motifs (top panel), only the N-terminal capping motif (middle panel) and only the C-terminal capping motif (bottom panel). Each capping arrangement includes different numbers of alternating consensus repeats “A” and “B”.

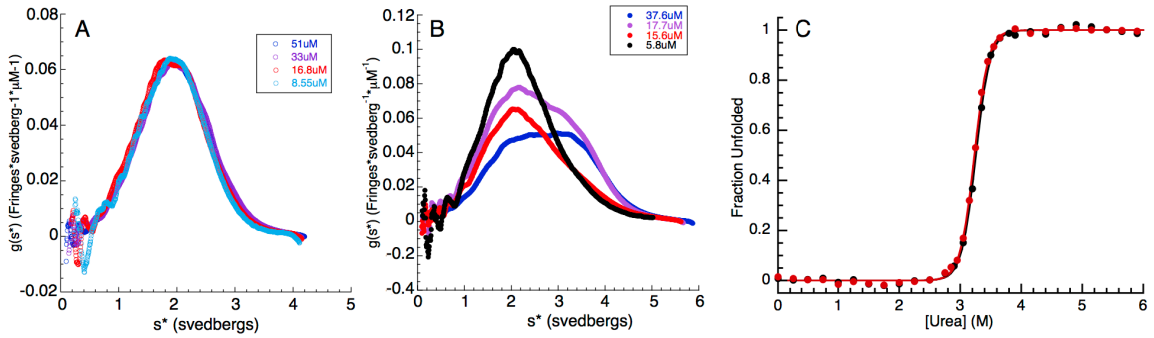


Figure 4.2. Concentration dependence of sedimentation and urea-induced denaturation of consensus LRR constructs. (A) g(s*) plot from sedimentation velocity AUC at four different concentrations for construct $N(AB)_2A$. Over the concentrations tested, $N(AB)_2A$ appears to be monomeric with a fitted s* value of 2.015, consistent with the size of the construct. (B) g(s*) plot from sedimentation velocity AUC at four different concentrations for construct $N(AB)_3C$, which appears to be monomeric at lower protein concentrations. (C) Overlay of urea-induced denaturation curves for $N(AB)_3C$ at 1 μM (black circles, black line) and 2 μM (red circles, red line). Lines results from fitting a two-state unfolding model to the data.

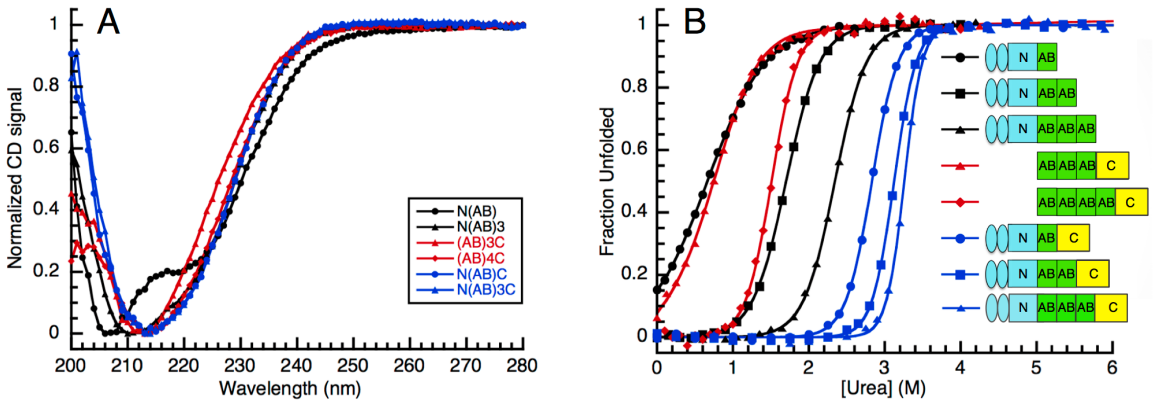


Figure 4.3. Far-UV scans and equilibrium unfolding of different consensus LRR constructs. (A) Far-UV CD spectra of representative constructs. The signals are normalized from zero to one. Spectra for different constructs have different shapes, depending on the compositions of secondary structures. (B) Urea-induced denaturation monitored by CD at 220 nm (filled circles). Lines result from fitting a two-state unfolding model to the data.

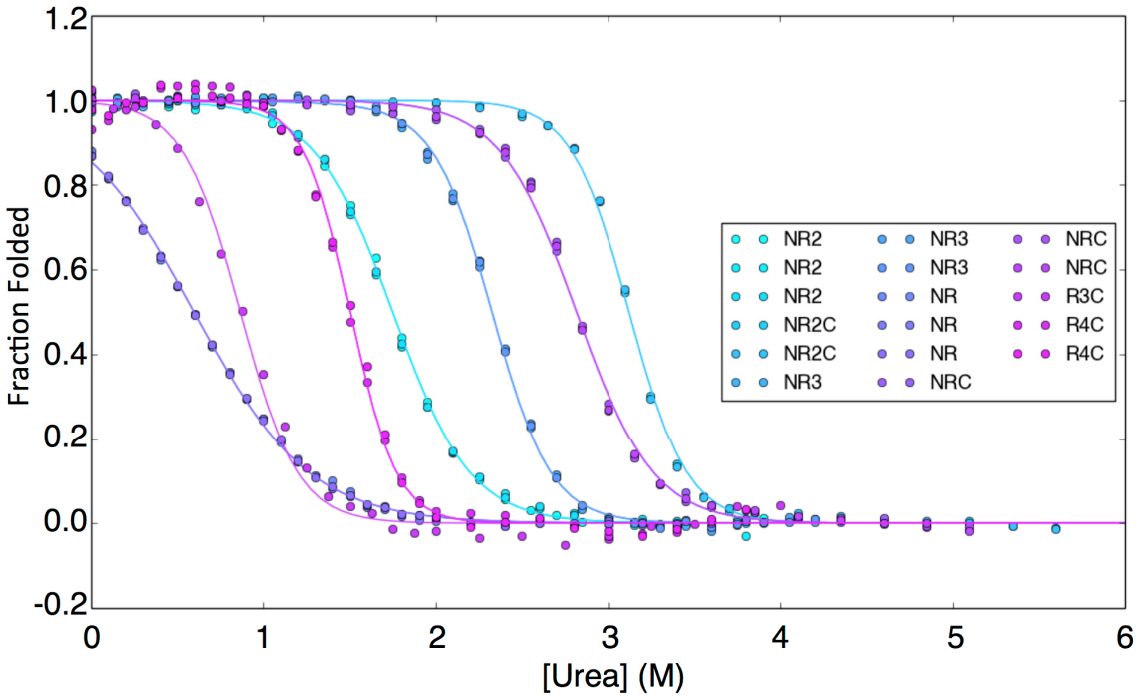


Figure 4.4. Ising analysis of unfolding transitions of consensus LRR constructs. Urea-induced denaturation monitored by CD at 220 nm (filled circle, continuous line). Lines result from global fits of an Ising model to the transitions.